



Community Long-term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS) Science Application Guides

Nadia Smith*
Rebekah Esmaili
Chris D. Barnet

*nadias@stcnet.com

Science and Technology Corporation
21 Enterprise Parkway, Suite 150, Hampton, VA, 23666

13 April 2021

Original Revision

CLIMCAPS V2 algorithm development was funded by NASA Grant No. 80NSSC18K0975

Original revision of the CLIMCAPS Science Applications User Guide was funded by the Atmospheric Infrared Sounder Project and the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Table of Contents

CHAPTER 1: ALGORITHM FLOW DIAGRAM	1-3
CHAPTER 2: GEOPHYSICAL RETRIEVAL VARIABLES.....	2-7
SECTION 1: TEMPERATURE	2-7
1. How can I access CLIMCAPS temperature retrievals?.....	2-7
2. Can I use CLIMCAPS temperature retrievals for studying climate trends?.....	2-7
3. Combined IR and MW temperature retrieval.....	2-7
4. MW-only Temperature Retrieval.....	2-11
5. Boundary Layer Adjustment	2-11
6. Temperature a-priori.....	2-11
7. Consider using CLIMCAPS temperature products for these applications.....	2-13
8. Relevant CLIMCAPS temperature product field names.....	2-13
SECTION 2: WATER VAPOR	2-14
1. How can I access CLIMCAPS water vapor retrievals?	2-14
2. Is there a way to quickly identify wet/dry regions for my application?	2-15
3. Which retrievals should I avoid?	2-15
4. Combined IR and MW H ₂ O retrieval	2-15
5. MW-only H ₂ O vapor retrieval	2-20
6. Boundary layer adjustment	2-21
7. H ₂ O vapor a-priori	2-21
8. Retrieval Evaluation	2-23
9. Units of conversion.....	2-25
10. Relevant fields in file.....	2-25
SECTION 3: OZONE.....	2-26
1. How can I access CLIMCAPS ozone retrievals?.....	2-26
2. Can I use CLIMCAPS products to study stratospheric ozone?.....	2-26
3. Can I use CLIMCAPS products to study tropospheric ozone?	2-27
4. CLIMCAPS algorithm flow with respect to O ₃	2-27
5. O ₃ information content and retrievals.....	2-31
6. Ozone a-priori	2-35
7. CLIMCAPS product field names relevant to O ₃ applications.....	2-36
SECTION 4: CARBON MONOXIDE	2-37
1. How can I access CLIMCAPS CO retrievals?	2-38
2. Which retrievals should I avoid?	2-38
3. How do I interpret CO retrievals in cloudy or smoky regions?	2-38
4. Which type of applications are CLIMCAPS CO good for?.....	2-38
5. Boundary layer adjustment	2-43
6. CO a-priori	2-43
7. Convolution of a 'truth' profile to CLIMCAPS CO retrieval.....	2-47
8. CLIMCAPS product field names relevant to CO applications.....	2-47
SECTION 5: CARBON DIOXIDE	2-48
1. How can I access CLIMCAPS CO ₂ retrievals?	2-48
2. Can I use CLIMCAPS CO ₂ retrievals for studying climate trends?	2-49
3. Sensitivity of the infrared spectrum to CO ₂	2-49
4. Regularization of the CO ₂ retrieval.....	2-56
5. Preparing CO ₂ retrievals for applications.....	2-57
SECTION 6: METHANE.....	2-61
1. How can I access CLIMCAPS CH ₄ retrievals?	2-61
2. Which retrievals should I avoid?	2-61

3. How do I interpret the spatial variability of daily CH ₄ retrievals?	2–61
4. Which type of applications are CLIMCAPS CH ₄ good for?	2–62
5. Preparing CH ₄ retrievals for applications.....	2–68
6. CH ₄ a-priori	2–70
7. CLIMCAPS product field names relevant to CH ₄ applications.....	2–71
CHAPTER 3: CLIMCAPS LEVELS AND LAYERS	3–76
SECTION 1: LEVELS VERSUS LAYERS.....	3–76
SECTION 2: CONVERTING LEVELS TO LAYERS	3–77
SECTION 3: CONVERTING LAYERS TO LEVELS	3–77
SECTION 4: BOUNDARY LAYER ADJUSTMENT.....	3–78
1. Pseudocode for adjusting boundary layer values.....	3–80
CHAPTER 4: CLOUD CLEARING AND CLOUD RETRIEVALS	4–82
1. Does CLIMCAPS retrieve cloud parameters?.....	4–82
2. Does CLIMCAPS retrieve soundings from clear-sky scenes only?	4–82
3. What is cloud clearing?	4–82
4. Does cloud clearing impact measurement information content?	4–82
SECTION 2: BRIEF SUMMARY OF METHODS	4–82
1. Relevant CLIMCAPS cloud and cloud clearing product fields	4–85
CHAPTER 5: WORKING WITH CLIMCAPS AVERAGING KERNELS	5–89
SECTION 1: INTERPRETING CLIMCAPS AVERAGING KERNELS	5–89
SECTION 2: TRAPEZOID STATE FUNCTIONS	5–97
SECTION 3: CONVOLVING WITH A REFERENCE PROFILE	5–99
1. What to expect when you convolve a reference profile.....	5–101
2. Constructing the transformation matrix and its inverse.....	5–102
3. IDL Routines.....	5–103
CHAPTER 6: QUALITY CONTROL.....	6–108
SECTION 1: VERTICAL QUALITY CONTROL FLAGS	6–108
SECTION 2: FOOTPRINT QUALITY CONTROL	6–110
1. Data filtering options	6–112
CHAPTER 7: CLIMCAPS INFRARED CHANNEL SELECTION.....	7–115
1. CLIMCAPS channel subsets for AIRS and CrIS.....	7–115
2. CLIMCAPS channel selection method	7–125
CHAPTER 8: USEFUL CLIMCAPS TABLES	8–127
APPENDIX B: LIST OF FIGURES	8–130
APPENDIX B: LIST OF TABLES	8–139

Chapter 1: Algorithm Flow Diagram

CLIMCAPS retrieves multiple Earth surface, cloud and atmospheric state variables from infrared and microwave measurements using Optimal Estimation inversion (Rodgers, 2000). We describe CLIMCAPS in detail elsewhere, with specific reference to uncertainty (Smith and Barnett, 2019) and information content (Smith and Barnett, 2020). Figure 1 outlines the main retrieval steps. Each box with a variable name represents an Optimal Estimation retrieval.

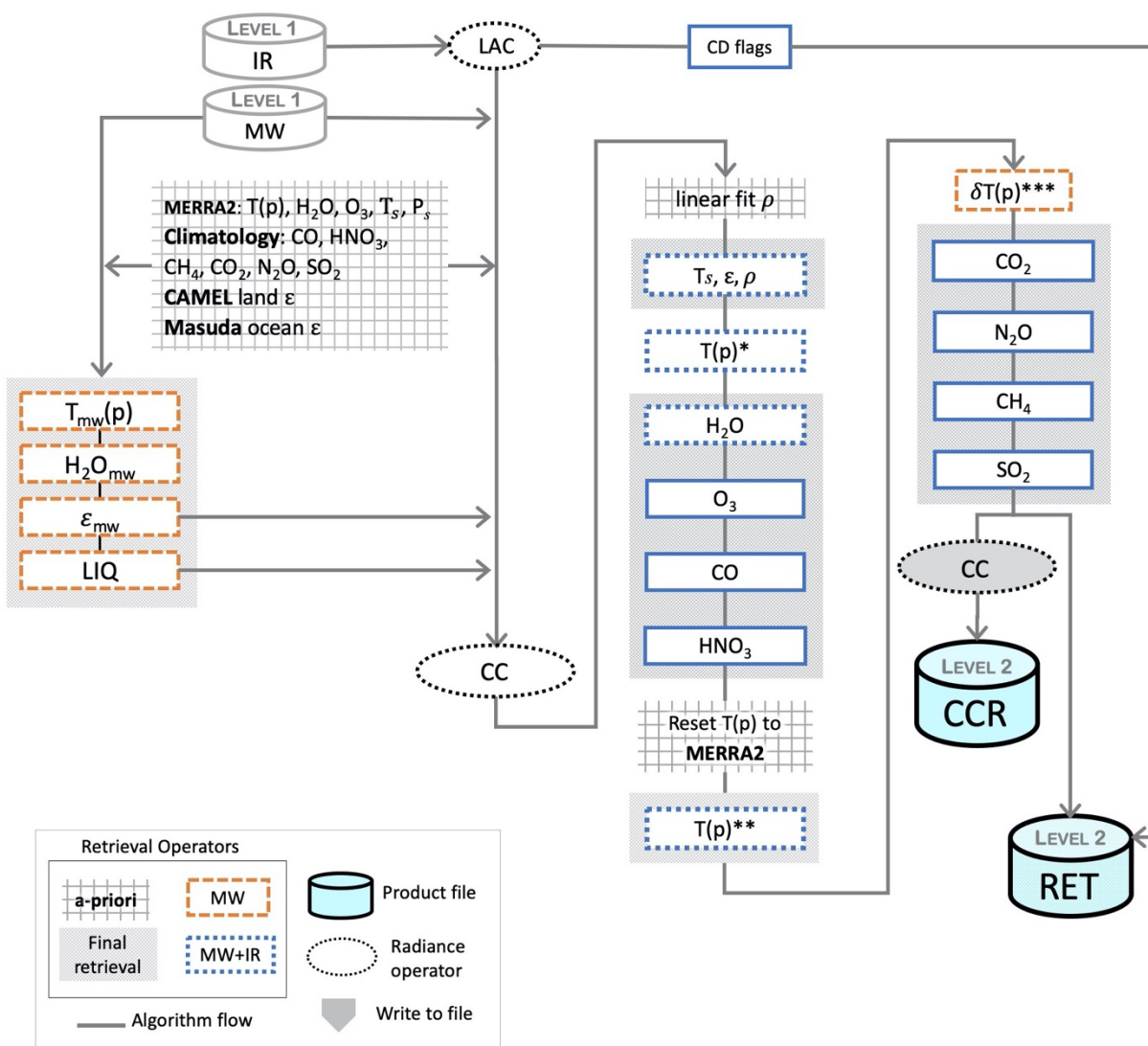


Figure 1: Flow diagram of the CLIMCAPS sequential retrieval algorithm. This gives a broad overview of the main retrieval steps and their logical flow towards two final products files, CLIMCAPS retrievals (CLIMCAPS RET) and cloud cleared radiances (CLIMCAPS CCR). Note that we discuss different aspects of CLIMCAPS algorithm flow also in (Smith and Barnett, 2019, 2020). See Table 1 for a description of the acronyms and symbols used here.

Table 1: List of acronyms and symbols used in Figure 1.

Table symbol	Description
ε	Earth surface emissivity
ρ	Earth surface reflectivity
$\delta T(p)^{***}$	Delta temperature: $T(p)$ with subset of MW channels minus $T(p)^{**}$ from MW+IR channels. This step tests the quality of the $T(p)^{**}$ retrieval.
CAMEL	CLIMCAPS implementation of the Combined ASTER and MODIS Emissivity database over Land (Hook, 2019c, a, b)
CC	Cloud Clearing that includes retrieval of cloud fraction and cloud top pressure
CCR	Level 2 Cloud Cleared Radiance product
CD flags	Constituent Detection flags for isoprene, ethane, propylene and ammonia
CH ₄	Methane retrieval on 100 pressure layers using a subset of IR channels
Climatology	Global representation of atmospheric variables
CO	Carbon monoxide retrieval on 100 pressure layers using a subset of IR channels
CO ₂	Carbon dioxide retrieval on 100 pressure layers using a subset of IR channels
H ₂ O	Water vapor retrieval on 100 pressure layers using a subset of MW+IR channels
HNO ₃	Nitric acid retrieval on 100 pressure layers using a subset of IR channels
IR	Infrared
LAC	Local angle correction of IR radiances within 3 x 3 fields of view
Level 1	NASA geolocated, calibrated radiance products for IR and MW measurements
Level 2	NASA geophysical products retrieved from Level 1 radiance measurements.
LIQ	Liquid water path
Masuda	CLIMCAPS implementation of the Infrared sea surface emissivity model: (Masuda et al., 1988) as modified by (Wu and Smith, 1997)
MERRA2	Modern-Era Retrospective analysis for Research and Applications Version 2 (GMAO, 2015) collocated in time and space to the CLIMCAPS instrument footprints.
MW	Microwave
N ₂ O	Nitrous oxide retrieval on 100 pressure layers using a subset of IR channels
O ₃	Ozone retrieval on 100 pressure layers using a subset of IR channels
P _s	Surface pressure
RET	Level 2 geophysical retrieval product
SO ₂	Sulphur dioxide retrieval on 100 pressure layers using a subset of IR channels
$T(p)^*$	Temperature retrieval on 100 pressure levels using a subset of MW+IR channels – first retrieval
$T(p)^{**}$	Temperature retrieval on 100 pressure levels using a subset of MW+IR channels – second and final retrieval
T _s	Surface skin temperature

The flowchart in Figure 1 outlines the main CLIMCAPS retrieval steps. We highlight only those steps that result in the primary retrieval products, namely cloud cleared radiances (CCR) and the nine profile retrievals – $T(p)$, H₂O, O₃, CO, CO₂, CH₄, SO₂, HNO₃, N₂O. CLIMCAPS profiles

are available on 100 pressure coordinates, with temperature (**air_temp**) represented on 100 pressure levels (**air_pres**) and the gas species on 100 pressure layers (**air_pres_lay**). For the sake of simplicity, we omit reference to those steps where (i) diagnostic metrics and quality control indices are calculated, and (ii) quantities are derived from the main retrieval variables, such as outgoing long-wave radiation (OLR) and relative humidity. CLIMCAPS output variables are saved to two product files in netCDF format:

- ***L2_CLIMCAPS_CCR.***: Cloud cleared radiances.
- ***L2_CLIMCAPS_RET.***: Retrieved and derived geophysical variables. This file is organized into two tiers, with the main group of variables as tier 1 (T(p), surface and cloud variables, relative humidity, etc.) and the remaining retrieved variables, derived quantities and diagnostic metrics organized into four subgroups on tier 2. We refer to three of these subgroups in Figure 1:
 - **aux**: collection of auxiliary variables, diagnostic metrics as well as minor gas detection flags.
 - **mol_lay**: collection of the eight primary trace gas profile retrievals, H₂O, O₃, CO, CO₂, CH₄, SO₂, HNO₃, N₂O. With the exception of CO₂ that is retrieved as volume mixing ratio, all trace gases are retrieved as layer column densities. The averaging kernel matrices for all nine retrieval variables (eight gases and T(p)) are contained in the **ave_kern** subgroup.
 - **mw** collection of MW-only retrieval variables using the algorithm developed by (Rosenkranz, 2001, 2006)

CLIMCAPS uses several diagnostic metrics to derive a single, final quality control flag (0 = best, 1 = good, 2 = bad). Some reasons for rejection include:

- The footprint is covered with precipitating clouds such that T_{MW}(p) fails
- The retrieved cloud fraction exceeds 80%
- Large differences exist between the clear radiance estimate (from the MW-only step) and the cloud clear estimate (from the cloud clearing step)
- The RMS of the observed minus calculated brightness temperature exceeds 1.75 K for specific window channels.
- The boundary layer estimate of $\delta T(p)^{***} > 1.5$ K

None of the above checks alter the retrieved values, but instead are used to derive a final quality flag. CLIMCAPS, therefore, does not have quality control flags tailored for each individual retrieved variable, but instead applies the logic that if T(p) and H₂O fails, all subsequent trace gas retrievals also fails. In future, we may re-evaluate and change this.

CLIMCAPS does not use any of the MW channels in cloud clearing or the cloud parameter retrievals, namely cloud fraction and cloud top pressure.

The MW-only retrieval of liquid water path (LIQ) and emissivity (ϵ) are used to define the background atmospheric state in subsequent IR+MW retrievals, while the MW-only temperature and water vapor retrievals are written to the output RET file as is. The MW-only retrievals of temperature and water vapor are NOT used in subsequent IR+MW retrievals.

References

- GMAO: MERRA-2 inst3_3d_asm_Nv: 3d,3-Hourly, Instantaneous,Model-Level,Assimilation,Assimilated Meteorological Fields V5.12.4, <https://doi.org/10.5067/WWQSXQ8IVFW8>, 2015.
- Hook, S.: Combined ASTER and MODIS Emissivity database over Land (CAMEL) Coefficient Monthly Global 0.05Deg V002, <https://doi.org/10.5067/MEASURES/LSTE/CAM5K30CF.002>, 2019a.
- Hook, S.: Combined ASTER and MODIS Emissivity database over Land (CAMEL) Emissivity Monthly Global 0.05Deg V002, <https://doi.org/10.5067/MEASURES/LSTE/CAM5K30EM.002>, 2019b.
- Hook, S.: Combined ASTER and MODIS Emissivity database over Land (CAMEL) Uncertainty Monthly Global 0.05Deg V002, <https://doi.org/10.5067/MEASURES/LSTE/CAM5K30UC.002>, 2019c.
- Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore; Hackensack, N.J., 2000.
- Rosenkranz, P. W.: Retrieval of temperature and moisture profiles from AMSU-A and AMSU-B measurements, 39, 2429–2435, <https://doi.org/10.1109/36.964979>, 2001.
- Rosenkranz, P. W.: Cloud liquid-water profile retrieval algorithm and validation, 111, <https://doi.org/10.1029/2005JD005832>, 2006.
- Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), 11, 1227, <https://doi.org/10.3390/rs11101227>, 2019.
- Smith, N. and Barnet, C. D.: CLIMCAPS Observing Capability for Temperature, Moisture and Trace Gases from AIRS/AMSU and CrIS/ATMS, Gases/Remote Sensing/Data Processing and Information Retrieval, <https://doi.org/10.5194/amt-2020-71>, 2020.
- Wu, X. and Smith, W. L.: Emissivity of rough sea surface for 8–13 μm : modeling and verification, 36, 2609, <https://doi.org/10.1364/AO.36.002609>, 1997.

Chapter 2: Geophysical Retrieval Variables

Section 1: Temperature

CLIMCAPS retrieves profiles of temperature in units [K] on a fixed vertical grid with 100 pressure levels. These are the standard pressure levels used in the stand-alone AIRS radiative transfer model, known as SARTA (Strow et al., 2003). We select a subset of channels for the CLIMCAPS temperature retrieval from the long-, mid- and short-wave infrared (IR) bands. Refer to the channel selection chapter for more details.

1. How can I access CLIMCAPS temperature retrievals?

CLIMCAPS T(p) retrievals are part of the main Level 2 product file that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://disc.gsfc.nasa.gov/>). There are temperature retrievals at every cloud-cleared retrieval scene (~50 km at nadir and ~150 km at edge-of-scan), twice a day from each instrument ascending and descending orbit.

2. Can I use CLIMCAPS temperature retrievals for studying climate trends?

We caution against using satellite sounding retrievals in calculating long-term trends without careful consideration for the influence from their a-priori and systematic sources of uncertainty in the measurements. In this chapter, we focus on the fact that T(p) retrievals depend on spectral channels primarily sensitive to CO₂ emissions. This means that the retrieval of long-term trends in T(p) – on the order of ~0.1 K per decade – depends on accurate knowledge of CO₂ decadal patterns, which is difficult to know globally. In turn, the retrieval of CO₂ mixing ratio strongly depends on knowledge of T(p) at every scene (See Section 1).

In CLIMCAPS V2 we take a different approach to AIRS V7 by having a reanalysis model, MERRA-2, as the a-priori for T(p). AIRS V7 uses a non-linear statistical regression (Milstein and Blackwell, 2016; Susskind et al., 2014) as the T(p) a-priori that is calculated at run-time using all IR channels, with the effect that the a-priori at each footprint is independent of its neighbors. There is, thus, no a-priori spatial structure. MERRA-2 T(p), on the other hand, has strong spatial structure and meso-scale gradients for T(p) in the troposphere and stratospheres. With a MERRA-2 as the T(p) a-priori, the CLIMCAPS T(p) retrieval inherits this spatial, temporal and vertical structure from its MERRA2 a-priori.

3. Combined IR and MW temperature retrieval

There are two chapters relevant to our discussion here; CLIMCAPS flow diagram as well as the CLIMCAPS water vapor (H₂O) retrieval. Similar to H₂O, CLIMCAPS retrieves T(p) in two stages; first with a set of microwave-only (MW) channels (**mw_air_temp**) using the method discussed in (Rosenkranz, 2001, 2003) and then with a combination of microwave and IR channels (MW+IR) using the method discussed in (Smith and Barnet, 2019, 2020). The MW+IR T(p) retrieval itself has two steps, as illustrated in the flow diagram. Note that only the final MW+IR retrieval step is written to the product file as **air_temp**.

We explain this two-step MW+IR retrieval approach in Smith and Barnet (2019, 2020) but can briefly summarize the two primary reasons. (1) The retrieval of H₂O and trace gas species from

an IR measurement depends on knowledge of temperature at the target scene. In CLIMCAPS, we retrieve $T(p)$ first, followed by H_2O and trace gases species in the order as depicted in the flow diagram. Having an estimate of $T(p)$ that is consistent with the MW+IR measurement ensures robust trace gas retrievals. Once we can account for scene-specific trace gases, we retrieve $T(p)$ a final time. (2) The CLIMCAPS $T(p)$ retrieval is useful for performing internal quality checks on the final product. For example, one quality check tests if the final retrieved temperature near the surface deviates significantly from an MW-only estimate of temperature. If the root mean square is > 1.5 K, then the retrieval is rejected even if all other checks were successful. This test does not change the state of the retrieved values, but instead identifies and flags failed retrievals.

The MW+IR $T(p)$ retrieval uses a subset of channels. We documented the subset of IR channels we selected for $T(p)$ in the channel selection chapter for the first and second retrieval steps. As far as the MW channels go, they vary between retrieval stages and instruments as detailed in Table 2.

Table 2: Channel selection from AMSU and ATMS for each of the two CLIMCAPS retrieval stages, microwave only (MW-only), and a combined MW and IR (MW+IR) retrieval. See CLIMCAPS flow diagram for methodology and IR channel selection chapter for the IR channel subsets.

Instrument (platform)	MW-only retrieval	MW+IR retrieval
	Channel number (total number of channels)	Channel number (total number of channels)
AMSU (Aqua)	3, 6, 8, 9, 10, 11, 12, 13, 14, 15 (10)	3, 6, 8, 9, 10, 11, 12, 13, 14 (9)
ATMS (SNPP, JPSS-1)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 (22)	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 (13)

The CLIMCAPS file contains a suite of diagnostic metrics with which to evaluate retrieval quality. Figure 2 below, shows the degrees of freedom (DOF) for $T(p)$ on 1 April 2016. The DOFs has a strong latitudinal dependence with highest values in the Tropics, with a maximum around ~ 3.6 .

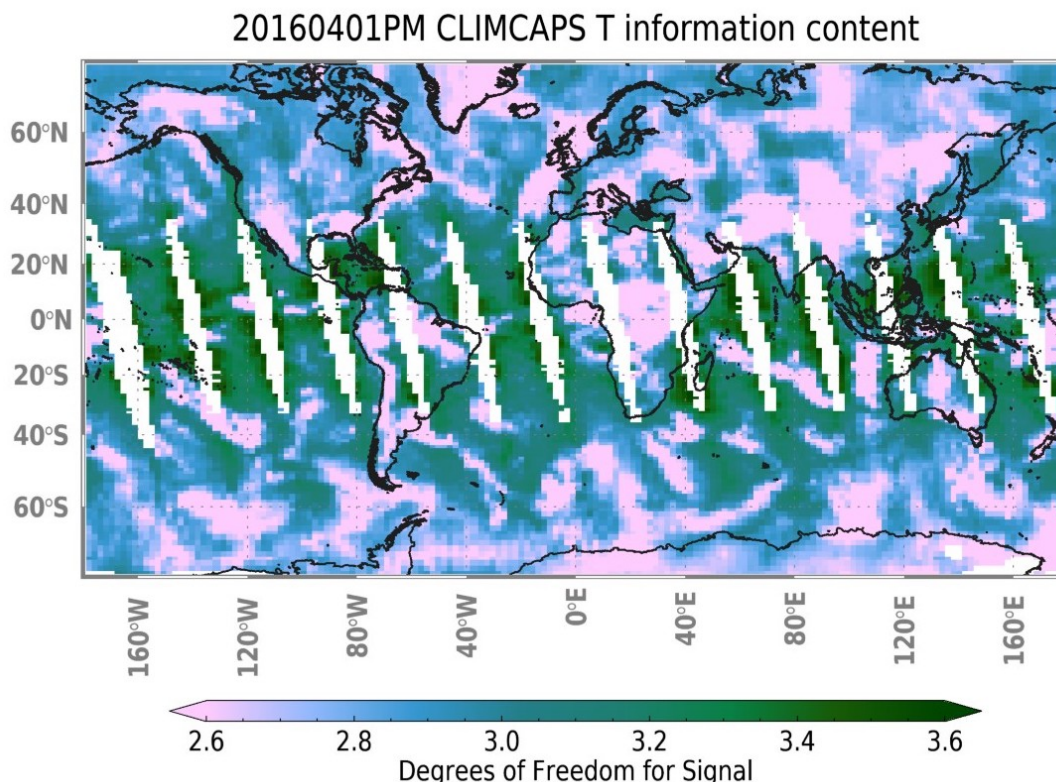


Figure 2: Information content as 'degrees of freedom' for CLIMCAPS-SNPP temperature (T) retrievals (**air_temp_dof**) from full spectral resolution cloud cleared CrIS radiances for an ascending orbit (13h30 local overpass time) as a global equal-angle grid on 1.5° resolution, close to single footprint size in the lower latitudes at edge of scan. We did not apply any quality control filtering (**air_temp_qc**) since the averaging kernels (**ave_kern/air_temp_ave_kern**) from which DOF is derived are unaffected by the quality of the retrieval. DOF, instead, characterizes the potential a sounding system has in retrieving a target variable (Smith and Barnett, 2020).

Figures 3 and 4 show the CLIMCAPS averaging kernels, temperature profile retrievals and their associated errors. We plot these for CLIMCAPS-SNPP on 1 April 2016. Figure 4 shows the mean profiles (with standard deviation error bars) for the Tropics (30°South to 30°North), and Figure 3 for the North Polar region (> 60°North). We used the diagonal vector of the averaging kernel matrix as the representation of the maximum sensitivity at each pressure level. The blue line (Figures 3 and 4, left panels) is the mean of the diagonal vectors in each latitudinal zone, respectively. Note how there are fewer vertical error bars on the blue line compared to the retrieval (orange line, center) and error (yellow line, right) profiles. This is because the 2-D averaging kernel matrices are written to the product file on the trapezoid pressure layers to save space. The retrieval and its error covariance matrix are, however, written out on the standard 100 pressure layers as 1-D arrays.

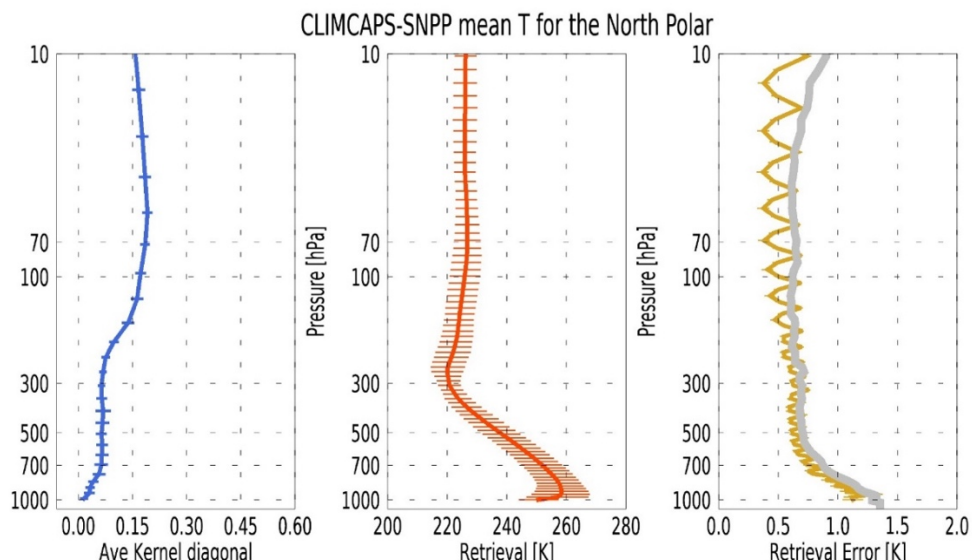


Figure 3: A diagnosis of CLIMCAPS-SNPP temperature retrievals for the North Polar latitudinal zone [$>60^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile and the error bars are the standard deviation at each pressure level. [left] CLIMCAPS temperature averaging kernel matrix diagonal vector from netCDF field `ave_kern/air_temp_ave_kern` that indicates the pressure levels at which CLIMCAPS has sensitivity to the true state of $T(p)$ in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field `air_temp` [K]. [right] CLIMCAPS retrieval error from netCDF field `air_temp_err` [K] represented here as percentage $[\text{air_temp_err}]/[\text{air_temp}] \times 100$. CLIMCAPS uses an empirical a-priori error estimate and is represented by the thick grey line. In addition, CLIMCAPS damps temperature by 20-25% with respect to MERRA-2 to improve the retrieval estimation of trace gases. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals. In calculating these mean profiles, we filtered out all retrievals where `air_temp_qc(*,i,j) ≥ 1` . We plot these profiles using the pressure level array from `air_pres*100` in hPa units.

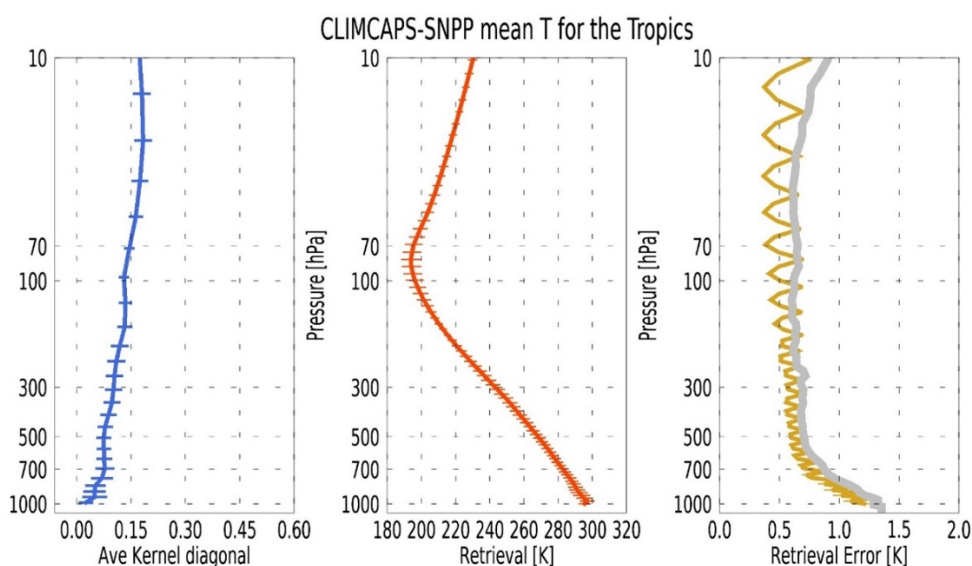


Figure 4: Same as Figure 2 but for the Tropical zone [30°S to 30°N].

The averaging kernel diagonal vectors for the North Polar zone (Figure 3) have less information content and vertical variability than those in the Tropics (Figure 4). This is consistent with the DOF shown in Figure 2, where information content of temperature is a strong function of latitude.

4. MW-only Temperature Retrieval

The CLIMCAPS system has a MW-only step that retrieves temperature (**mw_air_temp**), H₂O vapor column density (**mw_h2o_vap_mol_lay**), liquid water path (**mw_h2o_liq_mol_lay**), and surface emissivity (**surf_mw_emis**) using the method developed by (Rosenkranz, 2001, 2006), a sequential optimal estimation that uses a MW-only radiative transfer model as described by (Rosenkranz, 2003; Rosenkranz and Barnett, 2006). The liquid water path and surface emissivity retrieved variables are propagated into subsequent CLIMCAPS retrieval steps while temperature and H₂O vapor are written to the file as MW-only retrievals for use in research. Note that there are no corresponding averaging kernels for these MW-only retrievals.

MW-only estimates of temperature may be useful for certain applications where cloud clearing has failed due to uniform clouds or difficult surface conditions. However, the MW-only retrieval has a lower vertical resolution than the combined IR+MW CLIMCAPS retrieval, we caution against combining these in analyses without careful consideration.

5. Boundary Layer Adjustment

CLIMCAPS temperature uses a standard 100-level pressure grid to retrieve atmospheric variables from the Earth's surface to the top of atmosphere. This pressure grid is required by radiative transfer models (SARTA for CLIMCAPS) to accurately calculate top of atmosphere hyperspectral IR radiances. CLIMCAPS uses the exact same pressure grid at every scene on Earth and accounts for surface pressure as a separate variable during radiative transfer calculations. CLIMCAPS V2 uses MERRA2 surface pressure as input. The retrieved profiles are, however, reported on the 100-level grid as a means to standardize the output. It is important that you adjust the bottom level, i.e. that pressure level intersecting the Earth surface as identified by **air_pres_nsurf** in the CLIMCAPS netCDF file, to accurately reflect the T(p) at the surface. We refer the reader to the procedure described in Chapter 3, Section 4.

6. Temperature a-priori

CLIMCAPS employs MERRA2 (Gelaro et al., 2017; GMAO, 2015) as a-priori for its temperature, water vapor and ozone retrievals. CLIMCAPS converts MERRA2 temperature profiles from their 72 pressure levels to the standard 100 retrieval levels (**air_pres**). Additionally, CLIMCAPS interpolates MERRA2 profiles spatially and temporally to match the measurements. These interpolated MERRA2 profiles are written to the CLIMCAPS product file as **fg_air_temp**, **fg_h2o_mol_lay** and **fg_o3_mol_lay**.

We describe the benefits of employing MERRA2 as a-priori in Section 2.2.3 of Smith and Barnett (2019) and also explain how we derived the a-priori error covariance matrix depicted here. The reader should compare this section with the same one in the CLIMCAPS water vapor chapter that goes into more detail. Note that MERRA2, being a reanalysis model, has spatial correlation

in its temperature field, which will propagate into the CLIMCAPS temperature retrievals giving them smooth spatial gradients.

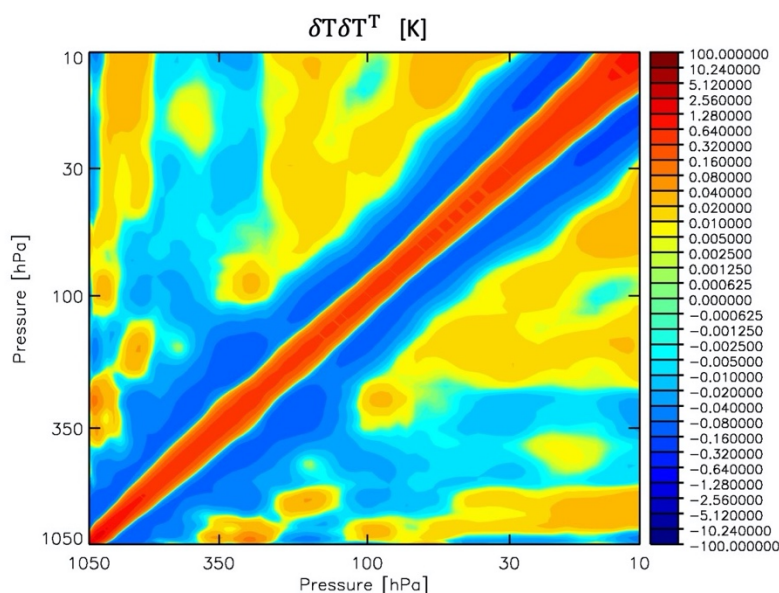


Figure 5: Empirical a-priori error covariance matrix used in CLIMCAPS V2 H₂O retrievals as described in Smith and Barnett (2019).

We depict the square root of the diagonal vector of the CLIMCAPS T(p) a-priori matrix (Figure 5) as the grey profile in Figures 3 and 4. Why, then, does the retrieval error profile (yellow in the same figures) have a zig-zag pattern? The simple answer is that this pattern emerges as a numerical artifact caused by our data compression methods.

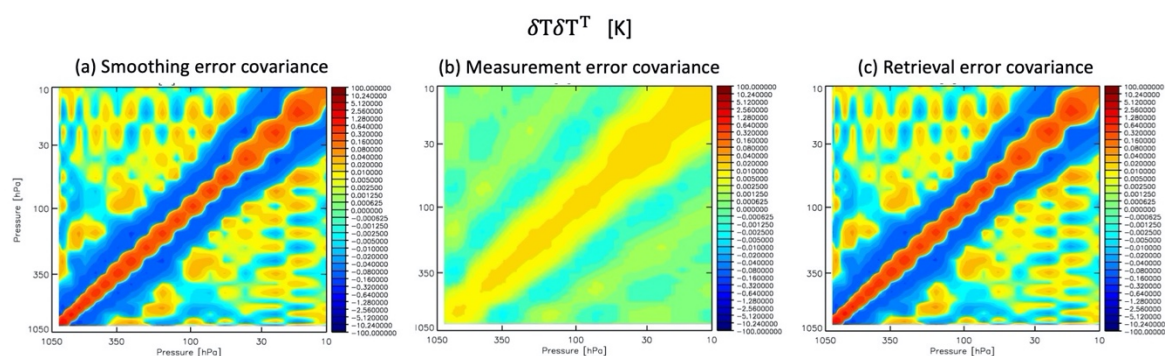


Figure 6: CLIMCAPS V2 smoothing error, measurement error and retrieval error covariance matrices as described in Smith and Barnett (2019).

The a-posteriori, or retrieval, error covariance matrix (Figure 6c) is the sum of the smoothing error (Figure 6a) and measurement error (Figure 6b) covariance matrices. The square root of the diagonal vector from the a-posteriori matrix (Figure 6c) is the yellow profile in Figures 3 and 4.

7. Consider using CLIMCAPS temperature products for these applications

CLIMCAPS temperature is a useful comparison to model results and useful for examining past weather events. For example, Figure 7 shows an example of a cold air outbreak over the Eastern United States that is associated with expansion of the polar vortex. When the jet stream weakens, cold arctic air can migrate southwards and cause below-average surface temperatures. These “polar vortex” events can occur frequently during the boreal winter (AAAS, 2001).

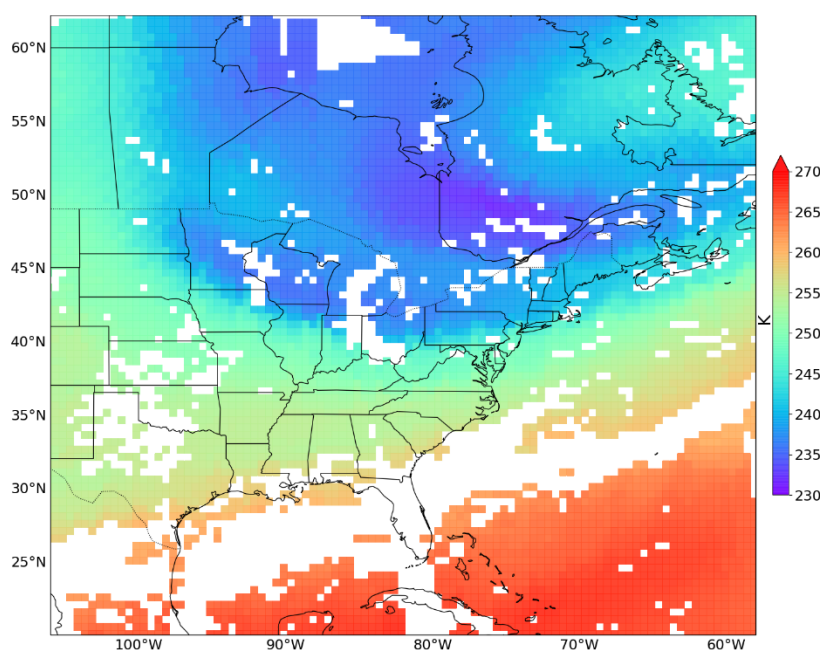


Figure 7: CLIMCAPS-SNPP T(p) at 500 hPa retrievals [K] from full-spectral resolution CrIS on 18 March 2019 from the ascending SNPP orbit. We filtered out all retrievals where `air_temp_qc(*,i,j) > 1`, which are shown as missing values.

8. Relevant CLIMCAPS temperature product field names

Within the netCDF files, the following fields are relevant for temperature studies. Each CLIMCAPS file contains 45 scanlines along track (atrack) and 30 FOR along each scanline, or across track (xtrack). With temperature profiles retrieved at each FOR on 100 pressure levels (air_pres), the arrays have dimensions [atrack, xtrack, airs_pres].

– Retrieved variables

- **air_temp**: MW+IR retrieved T(p) profile.
- **mw/mw_air_temp**: air temperature profile from the MW-only step.

– Retrieved surface variables

- **surf_temp**: MW+IR retrieved surface skin temperature.
- **mw/ mw_surf_temp**: MW-only retrieved surface skin temperature.
- **surf_ir_emis**: MW+IR retrieved IR surface emissivity.
- **mw/mw_surf_mw_emis**: MW-only retrieved surface emissivity.

- **surf_ir_refl**: Retrieved IR surface reflectivity.

– *Derived variables*

- **surf_air_temp**: near-surface temperature as retrieved MW+IR T(p) (air_temp) at surface pressure.
- **tpause_temp**: tropopause temperature as retrieved MW+IR T(p) (air_temp) at tropopause height.
- **mw/mw_surf_air_temp**: Near-surface air temperature as MW-only retrieved T(p) (mw_air_temp) at surface pressure.
- **cld_top_temp**: cloud top temperature derived as the temperature (air_temp) at cloud top pressure (cldfrac_tot).

– *Quality metrics*

- **Ave_kern/air_temp_ave_kern**: T(p) averaging kernel matrix.
- ***_dof**: The trace of the averaging kernel matrix as a measure of the number of pieces of information about the methane profile provided by the physical retrieval step. Degrees of freedom indicate the number of distinct vertical levels that the algorithm has sensitivity. For T(p), this is typically 3.6 or lower.
- ***_err**: Optimal-Estimation retrieval error, as the diagonal vector of the a-posterior error matrix.
- ***_qc**: profile quality control metrics ranging from 0 = good, 1 = suspect, 2 = bad.

– *A-priori variables*

- **fg_air_temp**: Air temperature profile from MERRA2 spatially, temporally and vertically interpolated to CLIMCAPS footprint and pressure grid (air_pres).
- **fg_surf_air_temp**: MERRA2 first guess for near-surface temperature.
- **fg_surf_temp**: MERRA2 first guess for surface skin temperature.
- **clim_surf_ir_refl**: a-priori IR surface reflectivity.

Section 2: Water Vapor

CLIMCAPS retrieves profiles of atmospheric water (H₂O) vapor column densities (molec/cm²) on the standard 100 pressure layers using subsets of spectral channels from both infrared (IR) and microwave (MW) measurements. The IR set includes channels from the longwave band, 780-1090 cm⁻¹ (~9.2–12.8 μm), and midwave band, 1213–1745 cm⁻¹ (~5.7–8.2 μm), with a total of 66 channels for CrIS and 46 for AIRS. The MW set includes channels from the 89 GHz line for AMSU and the 165.5 GHz and 183.31 GHz lines for ATMS.

1. How can I access CLIMCAPS water vapor retrievals?

CLIMCAPS H₂O retrievals are distributed within the main product file that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://disc.gsfc.nasa.gov/>).

2. Is there a way to quickly identify wet/dry regions for my application?

Yes. The CLIMCAPS netCDF file has a hierarchical structure to keep the fields organized according to type and application. The base level contains a few instrument and atmospheric variables that are easily accessible and frequently used in applications. Among them are three fields derived from the retrieved H₂O vapor column density. They are (i) **h2o_vap_tot**, or the total mass of water vapor content [kg/m²] more commonly referred to as total precipitable water (TPW) [mm] and, (ii) **rel_hum/spec_hum**, relative humidity [%] and specific humidity [kg/kg], respectively. Each of these fields have an associated quality control (qc) field, e.g., **h2o_vap_tot_qc** and you can use all values where qc is 0 or 1.

3. Which retrievals should I avoid?

The short answer is that you should avoid all retrievals where the quality control (qc) associated with a target field, e.g., **rel_hum_qc** for **rel_hum**, is equal to 2. There is a qc value for each level, so if the target field is a profile on pressure layers defined by **air_pres_h2o** (66 levels) then there will be 66 x qc values for that profile that flags the vertical layers either as 0 (best quality), 1 (good quality) or 2 (bad quality).

There are two unusual cases that may affect your application since they are not necessarily flagged as ‘bad quality’ (qc = 2).

– Cases of excessively dry layers

On rare occasions, CLIMCAPS H₂O retrieval profiles, specifically the netCDF fields **spec_hum**, **rel_hum**, and **mol_lay/h2o_vap_mol_lay**, have zero or unphysically low values in one or two pressure layers without being flagged as bad quality (qc = 2). These occur for two reasons: (1) due to MERRA2, the CLIMCAPS a-priori for H₂O, that contains extremely dry layers on the order of $\sim 10^{-20}$ kg/kg and, (2) due to numerical precision when the CLIMCAPS pre-processor converts these MERRA-2 profiles from their native 72 levels to the standard 100 retrieval layers (**air_pres_lay**) used in retrieval. We advise you to inspect these field and remove the entire profile where any layer value equals zero and your application is sensitive to this. **h2o_vap_tot** is not affected.

– Cases of super saturation

We added a quality control filter in CLIMCAPS that rejects retrievals (sets qc = 2) whenever relative humidity exceeds 110% anywhere between the Earth surface and 300 hPa. At pressure levels above 300 hPa, super saturation is physically common due to the combination of low H₂O concentrations and cold temperatures. This rejection criteria reduces the number of global retrievals by about 5% globally. Your application may require a more stringent humidity threshold, so we recommend that you evaluate results and create your own filter to better suit your application if necessary.

4. Combined IR and MW H₂O retrieval

CLIMCAPS uses spectral channels from both MW (AMSU/ATMS) and IR (AIRS/CrIS) measurements to retrieve H₂O. By combining MW and IR, CLIMCAPS can achieve and maintain robust retrieval quality in both clear and partly cloudy scenes. The contribution each set

of channels make to the final retrieval varies, since CLIMCAPS calculates scene-dependent information content and then weighs the spectral channels accordingly. For instance, clouds are opaque in the IR and thus one of the main sources of noise (Smith and Barnett, 2019). Relative to cloudy scenes, IR measurements of clear scenes have low scene-dependent noise. In clear cases, CLIMCAPS will thus give a strong weight to IR channels so that they make a larger contribution to the final retrieval. MW, on the other hand, has lower scene-dependent noise due to clouds (for all but non-precipitating clouds), so in complex cloudy scenes where IR information content is lower, CLIMCAPS will give stronger weight to the MW channels in its temperature and H₂O retrievals. CLIMCAPS thus leverages the different sensitivities and skills of IR and MW measurements to stabilize and optimize its retrievals.

Rather than using all available channels, CLIMCAPS use subsets of channels most sensitive to the target retrieval variable (e.g., H₂O) and with minimal sensitivity to other interfering variables. The CLIMCAPS H₂O channel set from IR measurements (AIRS and CrIS) is centered on water lines in the **midwave** – 3.7 μm , 6.6 μm – and **longwave** – 12 μm – spectral regions. For MW, CLIMCAPS uses channels in the 89 GHz band for AMSU and the 165.5 GHz as well as 183.31 GHz bands for ATMS. Despite careful channel selection, it is impossible to select channels that are sensitive only to a single target variable since absorption lines overlap, which means that the IR and MW channels each are sensitive to multiple atmospheric variables. For instance, in the 183 GHz band the O₃ and H₂O absorption lines overlap, and in the mid-wave IR, CH₄ and H₂O absorption features overlap. It is difficult to separate the highly mixed spectral measurements into distinct atmospheric variables and for this reason retrieval schemes typically start off with an a-priori (or background) estimate. The information in each channel subset are then extracted during the retrieval step and added to the a-priori state to change it according to the information available about the true state at a given point in space-time. CLIMCAPS uses MERRA2 as a-priori for its H₂O retrieval. This is a departure from the AIRS V5/V6/V7 retrieval approach that calculates a regression retrieval as a-priori. AIRS V5 employs a linear regression, while AIRS V6 and V7 uses a non-linear regression, or neural net. We refer the reader to Table 2 in Smith and Barnett (2019) where we compare and tabulate the differences between these two types of a-priori's and discuss the impacts they have on the final solution.

CLIMCAPS retrieves atmospheric state variables in sequence to optimize information content and stabilize the retrieval of multiple atmospheric variables from a single pair of IR and MW measured spectra (Smith and Barnett, 2019, 2020). The most linear variable is retrieved first, namely temperature, followed by H₂O and then the trace gases in specific order. Once CLIMCAPS retrieved H₂O, O₃, CO and HNO₃, and thus knows more about their concentrations at a specific scene, it retrieves temperature a second time to account for variability in the temperature spectral channel subset caused by these gases. The CLIMCAPS sequential retrieval approach, together with rigorous uncertainty quantification and propagation, promote a degree of independence among retrieved variables to minimize spectral correlation and improve their suitability for use in climate feedback studies (Dessler et al., 2008; Dessler and Wong, 2009; Smith and Barnett, 2019).

Knowledge of scene-dependent H₂O vapor helps stabilizes the retrieval of trace gases, not only in accounting for interfering H₂O absorption lines in the different channel sets, but also in calculating the displacement of dry air. With the exception of CO₂, CLIMCAPS retrieves trace

gases as molecules per cm^2 of *dry air*. Knowledge of H_2O vapor at a specific retrieval scene thus allows us to more accurately define the concentration of *dry air* in the a-priori and retrieval.

The MW+IR T(p) retrieval uses a subset of channels. We documented the subset of IR channels we selected for T(p) in the channel selection chapter for the first and second retrieval steps. As far as the MW channels go, they vary between retrieval stages and instruments as detailed in Table 3.

Table 3: Channels selected from AMSU and ATMS for each of the two CLIMCAPS retrieval stages, microwave (MW) only, and a combined MW and IR retrieval. See CLIMCAPS flow diagram for methodology and IR channel selection chapter for the IR channel subsets.

Instrument (platform)	MW-only stage	MW+IR stage
	Channel number (total)	Channel number (total)
AMSU (Aqua)	3, 6, 8, 9, 10, 11, 12, 13, 14, 15 (10)	15 (1)
ATMS (SNPP, JPSS-1)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 (22)	17, 18, 19, 20, 21, 22 (6)

The CLIMCAPS file contains a suite of diagnostic metrics with which to evaluate retrieval quality. In Figure 8 below, we plot the degrees of freedom (DOF) for H_2O on 1 April 2016 and note that it has a strong latitudinal dependence with highest values in the Tropics.

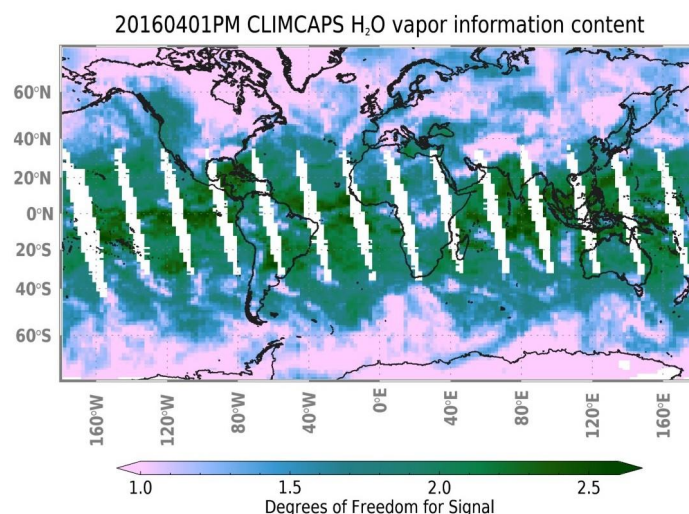


Figure 8: CLIMCAPS-SNPP degrees of freedom (DOF) for H_2O vapor at every retrieval scene from ascending orbits (01:30 PM local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about H_2O vapor at every scene. For most of the globe, CLIMCAPS has H_2O vapor DOF of ~ 1 . We used the netCDF field `h2o_vap_dof` and did not apply any quality filtering since DOF is not affected by retrieval outcome.

Figure 9 shows the CLIMCAPS-SNPP H₂O column density [molec/cm^2] (**mol_lay/h2o_vap_mol_lay**) and relative humidity [%] (**rel_hum**) for a global day on 1 April 2016. We applied quality filters and removed all retrievals where the quality control variables had values greater than one, i.e., where **mol_lay/h2o_vap_mol_lay_qc** > 1, or **rel_hum_qc** > 1. While both panels in Figure 9 are for the same global day and the same vertical pressure layer, ~850 hPa, you will notice large differences in their spatial patterns depending on the units. The retrieved quantity, H₂O vapor [molec/cm^2], has highest concentrations at low latitudes. The derived quantity, relative humidity [%], has high values across all latitudes even in atmospheres that are typically very dry such as the polar regions (> 60°). This is because the saturation vapor pressure is proportional to temperature, so a colder atmosphere can “hold” less H₂O vapor than a warmer atmosphere. Relative humidity thus describes the capacity of the atmosphere to hold water at a specific scene. Figure 9 illustrates how different units of measure can change the discussion surrounding atmospheric moisture. H₂O can be discussed in several other units of measure; such as mass mixing ratio, specific humidity, relative humidity, and Total Precipitable Water (TPW) elsewhere.

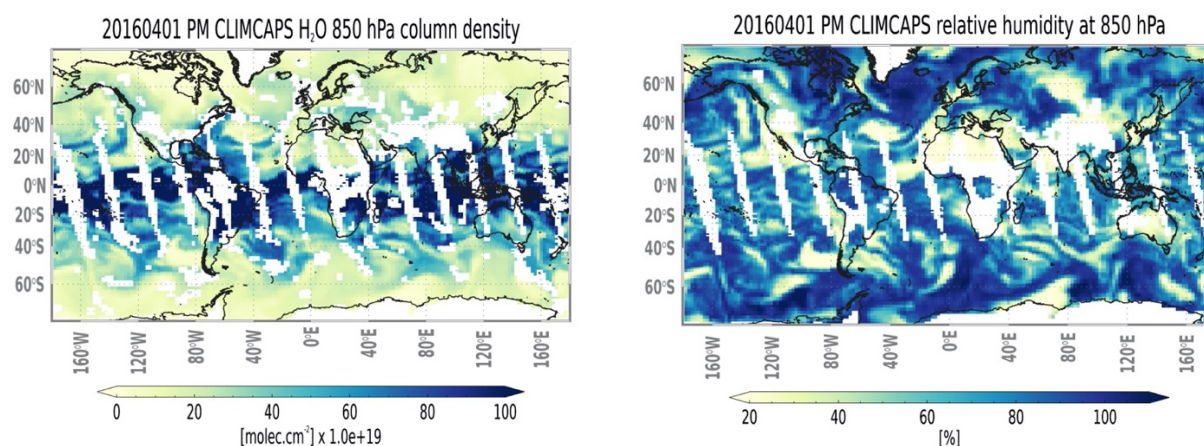


Figure 9: Global maps of CLIMCAPS-SNPP H₂O vapor fields in the lower troposphere around 850 hPa for (left) H₂O vapor column density [molec/cm^2] and (right) relative humidity [%]. H₂O column density is a retrieved variable and available in the netCDF file as **mol_lay/h2o_vap_mol_lay** on 100 pressure layers (**air_press_lay**). We selected values from layer 90 (839.98 hPa) and filtered out all retrievals where **mol_lay/h2o_vap_mol_lay_qc** > 1. Relative humidity is derived from **mol_lay/h2o_vap_mol_lay** and available in the netCDF file as **rel_hum** on 66 pressure layers (**air_press_h2o**). We selected values from layer 56 (852.79 hPa) and filtered out all retrievals where **rel_hum_qc** > 1.

In Figures 10 and 11 we take a closer look at CLIMCAPS averaging kernels, H₂O profile retrievals and their associated errors. We plot these for CLIMCAPS-SNPP on 1 April 2016. Figure 10 shows the mean profiles (with standard deviation error bars) for the Tropics (30°South to 30°North), and Figure 11 for the North Polar region (> 60°North). We used the diagonal vector of the averaging kernel matrix as representation of the maximum sensitivity at each pressure layer. The blue line is the mean of the diagonal vectors in each latitudinal zone, respectively. Note how there are fewer vertical error bars on the blue line compared to the retrieval (orange line) and error (yellow line) profiles. This is because the 2-D averaging kernel matrices are written to the product file on the trapezoid pressure layers to save space. The retrieval and its

error covariance matrix are, however, written out on the standard 100 pressure layers as 1-D arrays.

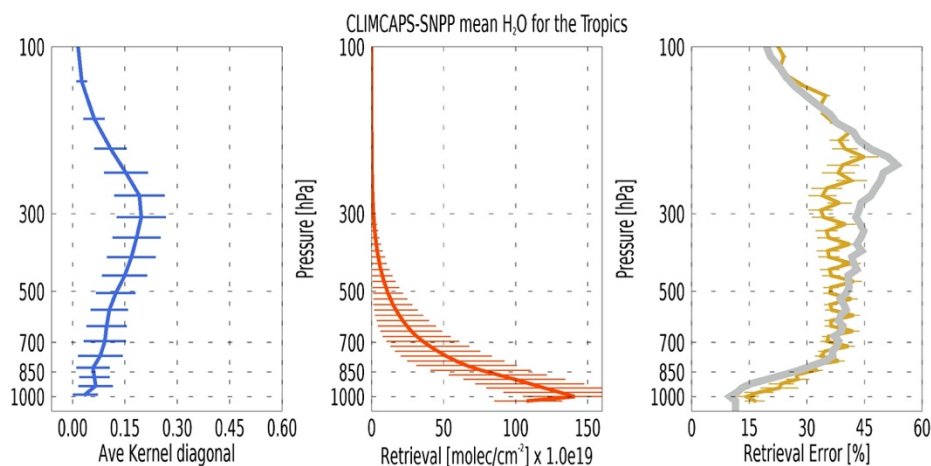


Figure 10: A diagnosis of CLIMCAPS-SNPP H₂O vapor retrievals in the Tropics [$>30^{\circ}\text{S}$, $<30^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile with error bars defined by the standard deviation at each pressure layer. [left] CLIMCAPS H₂O vapor averaging kernel matrix diagonal vector from netCDF field `ave_kern/h2o_vap_ave_kern` that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of H₂O vapor in the atmosphere. [middle] CLIMCAPS H₂O vapor profile retrieval from netCDF field `mol_lay/h2o_vap_mol_lay` [molec/cm²]. [right] CLIMCAPS retrieval error from netCDF field `mol_lay/h2o_vap_mol_lay_err` [molec/cm²] represented here as percentage $[\text{mol_lay/h2o_vap_mol_lay_err}]/[\text{mol_lay/h2o_vap_mol_lay}]*100$. CLIMCAPS uses an empirical a-priori error estimate and is represented by the thick grey line. In calculating these mean profiles, we filtered out all retrievals where `mol_lay/h2o_vap_mol_lay_qc(i,j) > 1`. We plot these profiles using the pressure layer array called `air_pres_lay`.

We calculated an empirical error covariance matrix for the CLIMCAPS H₂O vapor a-priori. The a-priori error profile in Figures 10 and 11 (righthand panel, grey line) is the square root of the diagonal vector from this error covariance matrix. In Section 2.2, we plot the full error covariance matrix and also explain why you see those zig-zag patterns in the retrieval error (righthand panel, yellow profile). A comparison between the retrieval error (yellow) and a-priori error (grey) profiles can be helpful in applications and data evaluation. As discussed elsewhere the retrieval error (`mol_lay/h2o_vap_mol_lay_err`) does *not* indicate retrieval accuracy, it does *not* quantify how closely the retrieval resemble the true state. Instead, the CLIMCAPS retrieval error is the a-posteriori error as described by (Rodgers, 2000) for Optimal Estimation (O-E) inversion systems. This a-posteriori error resembles the total error budget of the retrieval system and includes the measurement error, forward model error, background state error and smoothing error (Smith and Barnett, 2019). The a-posteriori error only has meaning in comparison to the a-priori error because it quantifies how much the measurements contributed to reducing uncertainty about the prior state. The amount of information extracted from the measurements depend on the inversion system (e.g., channel selection, measurement error covariance and various stabilization parameters) and is quantified by the averaging kernel.

When we re-examine Figures 10 and 11 we see that the a-posteriori error profile (righthand panel, yellow) is less than the a-priori error profile (righthand panel, grey) where the averaging kernel (left-hand panel, blue) reaches a maximum.

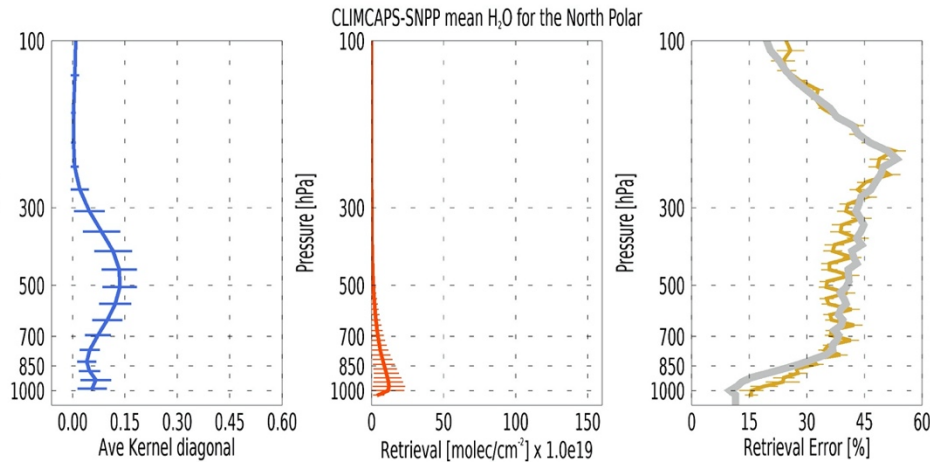


Figure 11: Same as Figure 3 but for the North Polar zone [$>60^{\circ}\text{N}$]

The averaging kernel diagonal vectors for the North Polar zone (Figure 11) are overall smaller and with a different shape than those in the Tropics (Figure 10). This is consistent with the DOF shown in Figure 8, where information content of H₂O is a strong function of latitude. Compared to the Tropics (Figure 10), retrievals in the North Polar zone (Figure 11) are an order of magnitude drier. In the Tropics, moisture extends high into the column since the ITCZ can produce deep convection, which can transport water molecules up to the tropopause and in some cases, into the stratosphere.

5. MW-only H₂O vapor retrieval

The CLIMCAPS system has a MW-only step that retrieves temperature (**mw_air_temp**), H₂O vapor column density (**mw_h2o_vap_mol_lay**), liquid water path (**mw_h2o_liq_mol_lay**), and surface emissivity (**surf_mw_emis**) using the method developed by (Rosenkranz, 2001, 2006), a sequential optimal estimation that uses a MW-only radiative transfer model as described by (Rosenkranz, 2003; Rosenkranz and Barnett, 2006). The liquid water path and surface emissivity retrieved variables are propagated into subsequent CLIMCAPS retrieval steps while temperature and H₂O vapor are written to the file as MW-only retrievals for use in research. Note that there are no corresponding averaging kernels for these MW-only retrievals.

MW-only estimates of H₂O vapor may be useful for certain applications where cloud clearing has failed due to uniform clouds or difficult surface conditions. However, the MW-only retrieval has a lower vertical resolution than the combined IR+MW CLIMCAPS retrieval, so we recommend caution when combining these in analyses.

6. Boundary layer adjustment

CLIMCAPS uses a standard 100-layer pressure grid to retrieve atmospheric variables from Earth surface to top of atmosphere. This pressure grid is required by radiative transfer models (SARTA for CLIMCAPS) to accurately calculate top of atmosphere hyperspectral IR radiances.

CLIMCAPS uses the exact same pressure grid at every scene on Earth and accounts for surface pressure as a separate variable during radiative transfer calculations. CLIMCAPS V2 uses MERRA2 surface pressure as input. The retrieved profiles are, however, reported on the 100-layer grid as a means to standardize the output. It is important that you adjust the bottom layer, i.e. that pressure layer intersecting the Earth surface as identified by **air_pres_lay_nsurf** in the CLIMCAPS netCDF file, to accurately reflect the total number of water vapor molecules in the boundary layer.

You should do this boundary layer adjustment if you calculate total column densities or if you are converting moisture units from layers to levels. We describe the method for doing this adjustment in Section 4 of Chapter 3.

7. H₂O vapor a-priori

CLIMCAPS employs MERRA2 (Gelaro et al., 2017; GMAO, 2015) as a-priori for its T(p), water vapor and ozone retrievals. CLIMCAPS converts MERRA2 T(p) profiles from their 72 pressure levels to the standard 100 retrieval layers (**air_pres_lay**). Additionally, CLIMCAPS interpolates MERRA2 profiles spatially and temporally to match the measurements. These interpolated MERRA2 profiles are written to the CLIMCAPS product file as **fg_air_temp**, **fg_h2o_mol_lay** and **fg_o3_mol_lay**.

We describe the benefits of employing MERRA2 as a-priori in Section 2.2.3 of Smith and Barnett (2019) and also explain how we derived the a-priori error covariance matrix depicted here. In short, the a-priori error (Figure 12) is presented as percentage because we derive $\delta(\text{H}_2\text{O})\delta(\text{H}_2\text{O})^T$ as the covariance of $(\text{ECMWF} - \text{MERRA2})/\text{ECMWF}$. This covariance matrix of percent difference is derived from an ensemble data set of 233,135 collocated profiles from ECMWF and MERRA2 for four global days, 1 January 2015, 1 April 2015, 1 July 2015 and 1 October 2015. Before calculating the difference between ECMWF and MERRA2, we apply a 3-point running mean on each profile individually to smooth out fine structures. Then we filter out all cases where the absolute value of the difference between ECMWF and MERRA2 is $> 50\%$ for H₂O and > 10 K for temperature.

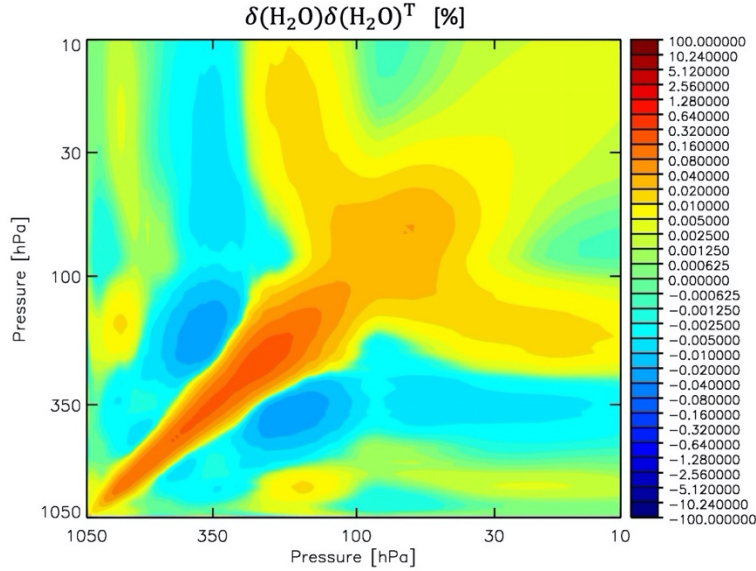


Figure 12: Empirical a-priori error covariance matrix used in CLIMCAPS V2 H₂O retrievals as described in Smith and Barnett (2019).

We depicted the square root of the diagonal vector of the CLIMCAPS H₂O a-priori matrix (Figure 12) as the grey profile in Figures 10 and 11. Why, then, does the retrieval error profile (yellow in the same figures) have a zig-zag pattern? The simple answer is that this pattern emerges as a numerical artifact caused by our data compression methods.

As described in Smith and Barnett (2020), CLIMCAPS calculates brute force Jacobians (weighting functions) on a reduced set of pressure layers as defined by overlapping trapezoid functions. We do this to speed up processing. In addition, CLIMCAPS performs its iterative minimization equation (i.e., retrieval) in orthogonal space as a set of eigenvectors to separate signal from noise. Every time the forward model, SARTA, is called, CLIMCAPS has to reconstruct the profiles from orthogonal to physical space and then convert them back onto the 100 layers (**air_pres_lay**). This numerical process introduces the oscillating pattern. For example, when CLIMCAPS ingests the matrix in Figure 12 as a-priori error covariance, transform it onto trapezoid layers, compress it to orthogonal space and then reconstruct and expand again according to Eq. 1, the matrix structure changes as seen in Figure 13a.

$$(\delta(\text{H}_2\text{O})\delta(\text{H}_2\text{O})^T)_a = G * \delta(\text{H}_2\text{O})\delta(\text{H}_2\text{O})^T * G \quad 1$$

Where,

$$G = F_{Lj} * U_{jk} * (I - \phi) * U_{kj} * F_{Lj}^{-1},$$

U_{jk} is the transformation matrix from k eigenvectors to j trapezoidal layers

F_{Lj} the transformation matrix from j to L SARTA pressure layers.

$k \leq 6$ eigenvectors

$j = 22$ trapezoid layers

$L = 100$ retrieval layers

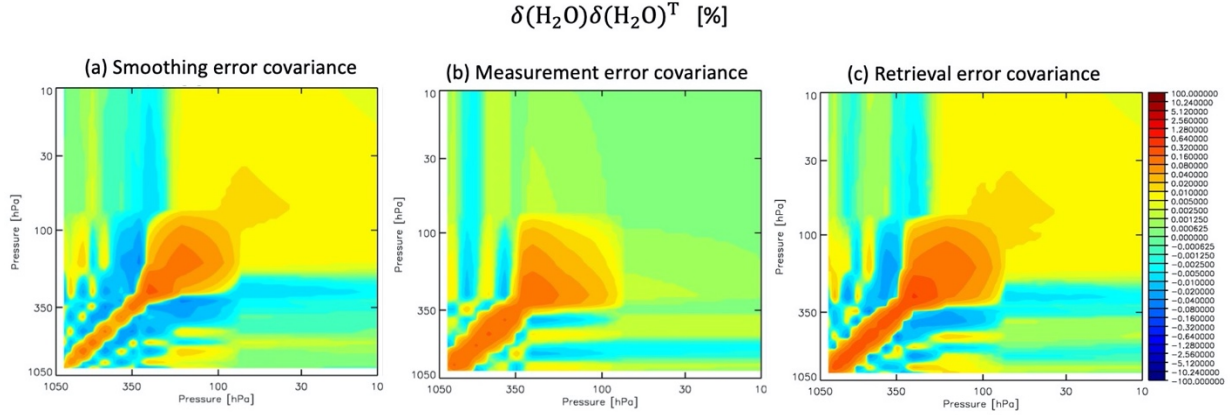


Figure 13: CLIMCAPS V2 smoothing error, measurement error and retrieval error covariance matrices as described in Smith and Barnett (2019).

The a-posteriori, or retrieval, error covariance matrix (Figure 13c) is the sum of the smoothing error (Figure 13a) and measurement error (Figure 13b) covariance matrices. The latter is calculated as:

$$(\delta(\text{H}_2\text{O})\delta(\text{H}_2\text{O})^T)_o = \mathbf{Z} * 1/\lambda \mathbf{Z}^T \quad 2$$

Where λ is the a-priori matrix, $\delta(\text{H}_2\text{O})\delta(\text{H}_2\text{O})^T$ in orthogonal (or eigenvector) space, k , and \mathbf{Z} is the amount of measurement (or measurement noise) that is believed in retrieval space (L) such that $\mathbf{Z} = \mathbf{F} * \mathbf{U}_{jk} * \phi$, where ϕ is the fraction of the parameter solved for. The square root of the diagonal vector from the a-posteriori matrix (Figure 13c) is the yellow profile in Figures 10 and 11.

8. Retrieval Evaluation

– Radiosondes/Dropsondes

Figure 14 below shows the atmospheric structure captured by dropsondes and CLIMCAPS retrievals respectively. The middle and bottom panels are the profiles along a NOAA research flight path, which passed over Hurricane Jerry on Sept 18, 2019 in the Atlantic Ocean. The flights were timed to coincide with the satellite overpass and are all within 2 hours of the flight. The dropsondes, unlike radiosondes, take approximately 20 minutes to reach the surface once they are released from the aircraft. They have less horizontal drift, typically around 50km. Dropsondes can sample over 10,000 pressure levels, which is two orders of magnitude higher than the 100 retrieval layers used in CLIMCAPS. Despite vertical and horizontal resolution differences, we see that both the dropsondes and CLIMCAPS observe the pockets of dry air. Knowledge of dry air is useful when examining Hurricanes since they are indicators of a storm's ability to intensify or weaken. In the case of Jerry, dry air from the Saharan Air Layer (SAL) was being entrained, but Jerry still briefly intensified to a category 3 Hurricane.

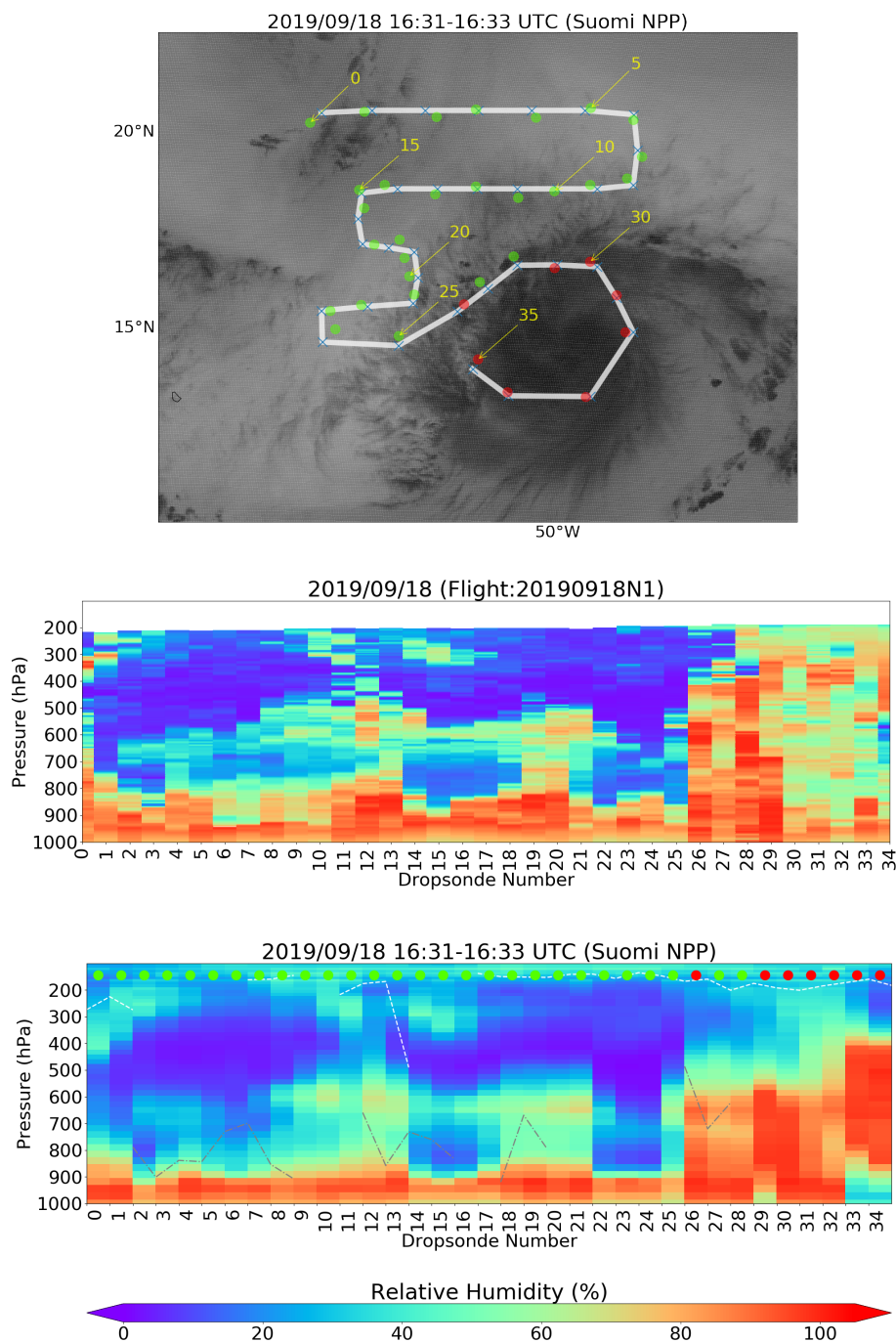


Figure 14: A research flight on September 18, 2019 shows (a) along a flight path over Hurricane Jerry the relative humidity profiles from (b) dropsondes released from a Gulfstream-IV "Hurricane Hunter" aircraft and (c) CLIMCAPS-SNPP H₂O retrievals as relative humidity. The solid line in (a) represents the flight path and the numbers are used to identify the location of the dropsonde profiles in (b) and (c). The colored dots indicate the center location of the CLIMCAPS footprint and if the retrieval passed (green) or failed (red). No averaging kernel convolution has been applied to the radiosonde data. However, for a quantitative comparison, we recommend that users apply this procedure.

We recommend work by (Nalli et al., 2018a, 2018b) that describe methods for inter-comparing satellite soundings and radiosondes.

– *Averaging Kernels*

Averaging kernels are helpful when evaluating and comparing satellite sounding retrievals as discussed in these publications (Iturbide-Sanchez et al., 2018; Maddy and Barnett, 2008; Smith and Barnett, 2020). We outline a method for convolving radiosondes (or model profiles) with CLIMCAPS averaging kernels to remove differences due to vertical resolution.

9. Units of conversion

Moisture is essential for measuring many derived atmospheric stability parameters. To facilitate these calculations, there are numerous ways to measure the water content of the atmosphere. CLIMCAPS retrieves moisture in column density with units of molec/m² (**mol_lay/h2o_vap_mol_lay**). In addition, the netCDF file contains a number of derived values, including relative humidity (**rel_hum**), specific humidity (**spec_hum**) and TPW (**h2o_vap_tot**). We describe unit conversions for H₂O elsewhere.

10. Relevant fields in file

– *Retrieved variables*

- **mol_lay/h2o_vap_mol_lay**: CLIMCAPS retrieval of H₂O vapor layer column densities [molec/m²] on 100 pressure layers (**air_pres_lay**). Multiply by 1.0e-04 to convert to the more common form of [molec/cm²]. measurements.
- **mw/mw_h2o_vap_mol_lay**: Microwave-only retrieval of H₂O vapor layer column densities [molec/m²] on 100 pressure layers (**air_pres_lay**). Multiply by 1.0e-04 to convert to the more common form of [molec/cm²].

– *Derived variables*

- **rel_hum**: Relative humidity [%] derived from **mol_lay/h2o_vap_mol_lay**, reported on 66 pressure layers (**air_pres_h2o**).
- **h2o_vap_tot**: Total mass of water vapor content [kg/m²] derived from **mol_lay/h2o_vap_mol_lay**. Also known as Total Precipitable Water, or TPW in units [mm], because 1 kg/m² = 1 mm. TPW refers to the gaseous form of water in the atmosphere, not water droplets in clouds.
- **spec_hum**: specific humidity [kg/kg], which is the mass fraction of H₂O in total air derived from **mol_lay/h2o_vap_mol_lay** and reported on 66 pressure layers (**air_pres_h2o**).
- **spec_hum_sat_ice**: saturation specific humidity over ice and reported on 66 pressure levels (**air_pres_h2o**) derived from **air_temp**.
- **spec_hum_sat_liq**: saturation specific humidity over liquid water and reported on 66 pressure levels (**air_pres_h2o**).
- **mw/mw_h2o_vap_tot**: Total mass of water vapor content [kg/m²] derived from **mw/mw_h2o_vap_mol_lay**. Also known as Total Precipitable Water, or TPW in units [mm], because 1 kg/m² = 1 mm. TPW refers to the gaseous form of water in the atmosphere, not water droplets in clouds.

- **mw/mw_spec_hum**: Specific humidity, the mass fraction of water vapor in total air, from the MW-only step [kg/kg] derived from **mw/mw_h2o_vap_mol_lay** and reported on 66 pressure layers (**air_pres_h2o**).

– *Quality metrics*

- **ave_kern/h2o_vap_ave_kern**: H₂O vapor averaging kernel matrix for every retrieval scene.
- **h2o_vap_dof**: The trace of the averaging kernel matrix as a measure of the number of independent pieces of information about H₂O at a target scene. For H₂O, CLIMCAPS typically maintain a DOF below 4, which means that CLIMCAPS has sensitivity to the true state of H₂O at a target scene in three vertical layers.
- ***_err**: The retrieval error estimate for column density (**h2o_vap_mol_lay**), relative humidity (**rel_hum**), specific humidity (**spec_hum**), and column total (**h2o_vap_tot**). The error has the same units as the original retrieved or derived variables.
- ***_qc**: The quality control flag for retrieved H₂O in several units of measure: column density (**h2o_vap_mol_lay**), relative humidity (**rel_hum**), specific humidity (**spec_hum**), and column total (**h2o_vap_tot**). This is the same quality flag used in the retrieved and derived variables.
- **aux/fg_h2o_vap_mol_lay**: MERRA2 a-priori for H₂O retrieval in the MW-only and MW+IR retrieval steps. These the native MERRA2 H₂O profiles converted to column densities [molec/m²] and interpolated spatially, temporally and vertically to match the CLIMCAPS retrieval footprints. This field is included in the netCDF file so that you can readily compare the retrieval with its a-priori.

Section 3: Ozone

CLIMCAPS retrieves profiles of ozone (O₃) layer densities as molecules per cm² [molec/cm²] on the standard 100 vertical pressure layers at each retrieval footprint, also known as ‘field of regard’. CLIMCAPS retrieves O₃ from a subset of hyperspectral infrared (IR) channels centered on the ~1030 cm⁻¹ (~9.7 μm) O₃ absorption line.

1. How can I access CLIMCAPS ozone retrievals?

CLIMCAPS O₃ retrievals are part of the main Level 2 product file that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <http://disc.gsfc.nasa.gov/>).

2. Can I use CLIMCAPS products to study stratospheric ozone?

Yes. CrIS and AIRS are passive instruments that make IR measurements of emitted radiance [mW/m²/steradian/cm⁻¹] at the top of atmosphere. This means their measurements are sensitive to the atmospheric state from the ‘top down’ with higher sensitivity to stratospheric emissions. For gases such as H₂O vapor, CO and CH₄ this holds no advantage because their stratospheric concentrations are negligible. Ozone, on the other hand, has stratospheric concentrations that far exceeds those from the troposphere so its IR spectral signature is distinct and strong (Figure 15b;

Figure 16). Stratospheric emission signals additionally benefit from low interference due to clouds, surface properties and pollutant gases. CLIMCAPS stratospheric O₃ retrievals from CrIS and AIRS measurements are, therefore, robust and can support stratospheric O₃ science and applications such as Gray (2019), Wargan et al. (2018) and Witze (2020).

3. Can I use CLIMCAPS products to study tropospheric ozone?

Sometimes. Unlike CO₂ that has different spectral signatures for stratospheric and tropospheric CO₂, O₃ from different parts of the atmosphere generate spectral signatures that overlap in the $\sim 1030\text{ cm}^{-1}$ region (Figure 15). Retrieving tropospheric O₃ (weak signal) means treating stratospheric O₃ (strong signal) as spectral interference. In the methane chapter, we discussed the challenge in retrieving weak signals (CH₄) from channels with strong interference (H₂O). CLIMCAPS can retrieve tropospheric O₃ with a degree of certainty when conditions allow, such as low stratospheric O₃, high temperature lapse rates, high surface temperature, low tropospheric H₂O vapor, absence of clouds and/or homogeneous surface emissivity. We diagnosed CLIMCAPS observing ability for tropospheric O₃ in Section 3.2 of Smith and Barnett (2020).

A number of diagnostic metrics in the CLIMCAPS product file support the inclusion of CLIMCAPS O₃ in data inter-comparisons and global evaluations such as Gaudel et al. (2018) and Cooper et al., (2014). We envisage that CLIMCAPS O₃ could also support severe weather research by characterizing stratospheric O₃ intrusion events (Berndt and Folmer, 2018; Dreessen, 2019; Langford et al., 2018). We advise against using CLIMCAPS O₃ in air quality monitoring since lower tropospheric information content is too low.

4. CLIMCAPS algorithm flow with respect to O₃

It is difficult to decompose (or invert) the information content of spectral measurements into distinct atmospheric variables suitable for science applications and climate studies. Once achieved for a single measurement under specific conditions, it is perhaps even more challenging to then replicate the same degree of robustness and quality in spectral inversions anywhere on the globe, day and night, Summer through Spring, year in, year out. Space-based IR and microwave (MW) measurements have highly correlated information about multiple atmospheric variables in each channel with signal-to-noise ratios changing from scene to scene. We demonstrate this for CO₂ and CH₄ elsewhere and for O₃ in Figures 15 and 16 below.

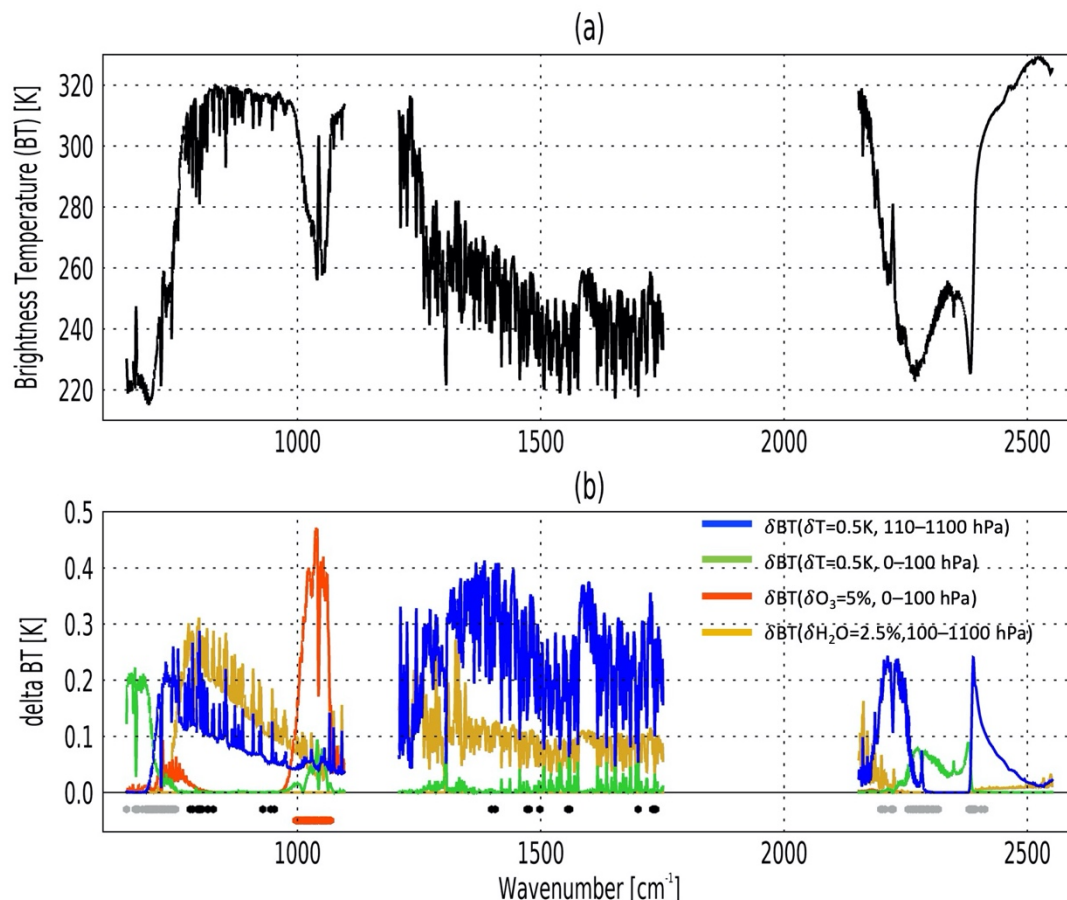


Figure 15: **(a)** Simulated CrIS spectra for the longwave (648.75–1096.25 cm⁻¹), mid-wave (1208.75–1751.25) and shortwave (2153.75–2551.25) bands using SARTA (Strow et al., 2003) with atmospheric state defined by the first CLIMCAPS V2 retrieval (scanline=1, footprint=1) of granule 104 on 1 April 2018 from an ascending orbit (13h30 local overpass time). SARTA simulates CrIS spectra in radiance units [mW/m²/steradian/cm⁻¹], which we converted to brightness temperature [K] at scene temperature. **(b)** Absolute values of Brightness Temperature differences (delta BT) to illustrate absorption features for temperature (T), ozone (O₃) and water (H₂O) vapor given a (blue) dT = 0.5 K perturbation in tropospheric T, 110–1100 hPa, (green) dT = 0.5 K perturbation in stratospheric T, 0–100 hPa, (red) dO₃ = 5% stratospheric O₃ perturbation, 0–100 hPa, and (gold) dH₂O = 2.5% tropospheric H₂O perturbation, 100–1100 hPa. Dots below the zero line indicate the CLIMCAPS retrieval channel sets for O₃ (red) as well as the first (grey) and second (black) pass of T retrievals.

Figure 15a is the full resolution CrIS spectrum in brightness temperature (BT) units [K] that we simulated with SARTA (Strow et al., 2003) using CLIMCAPS-SNPP retrievals (temperature (T), H₂O vapor, trace gases, surface temperature and cloud parameters) as state variables, and specifically those from scanline 1, footprint 1, ascending (daytime) granule 101 on 1 April 2018. Figure 15b depicts the spectral signatures for temperature, O₃ and H₂O in the long-, mid- and shortwave IR bands as the absolute value of the kernel functions, K, given by the change in BT, y, for a change in state variable, x, as:

$$K = \left| \frac{\delta y}{\delta x} \right| = |F(x) - F(x - p)| \quad 3$$

where F is the SARTA forward model with which we calculate a simulated CrIS radiance spectrum [$\text{mW/m}^2/\text{steradian/cm}^{-1}$] and convert to BT [K] at scene temperature, x is the state variable perturbed by the factor p . In Figure 15b, we perturbed tropospheric H_2O by 2.5%, thus $p = 0.025(x)$, stratospheric and tropospheric temperature respectively by 0.5 K, thus $p = 0.5$, and stratospheric O_3 by 5%, thus $p = 0.05(x)$. Kernels are typically 2-D matrices with dimensions $[m \times n]$, where m is the number of spectral channels and n the number of pressure layers. Our brute force kernels in Figure 15b have $n = 1$ because we perturb the target variable along a single broad layer for the sake of simplicity and illustration. In retrieval systems, $1 < n \leq 100$ to maximize information content and vertical resolution, given system constraints.

Of note in Figure 15b is the strength of the spectral kernel for a 2.5% perturbation in stratospheric O_3 [0–100 hPa] in the $\sim 1030 \text{ cm}^{-1}$ wavenumber region. It is from this region that we selected the set of channels for CLIMCAPS O_3 retrievals. A series of red dots below the zero line in Figure 15b indicates the spectral range of this channel set and we see that while these channels are dominated by stratospheric O_3 emission (red), they have interference signals from tropospheric H_2O vapor (gold) and temperature (blue) as well as stratospheric temperature (green). Additionally, the stratospheric and tropospheric spectral kernels for a 5% change in O_3 overlap in the $\sim 1030 \text{ cm}^{-1}$ region as depicted in Figure 16 below, which complicates their separation during retrieval.

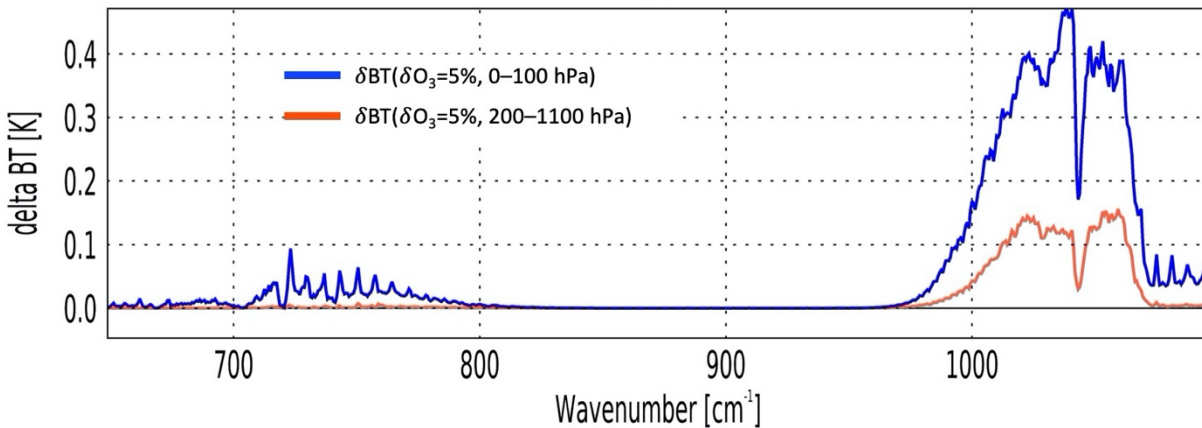


Figure 16: Same as Figure 1b but for O_3 kernels with 5% perturbation in the (blue) stratosphere [0–100 hPa] and (red) troposphere [200–1100 hPa].

There is a second stratospheric O_3 emission feature in the 700–800 cm^{-1} longwave range as seen in Figures 15b and 16. We do not use any channels from this region in the O_3 retrieval because it is buried within the CO_2 absorption region; however, knowledge of scene-dependent O_3 contributes to lower tropospheric retrievals of temperature. Note how the channel set used in CLIMCAPS temperature retrievals (grey + black dots below the zero line in Figure 15b), samples this stratospheric O_3 spectral feature specifically, which means that the CLIMCAPS temperature retrievals depend on knowledge of stratospheric O_3 to correctly account for

variability in these channels at a retrieval scene. We elaborate on this in the discussion of Figure 17 below.

We adopted a sequential retrieval approach in CLIMCAPS (Smith and Barnett, 2019, 2020) to stabilize the retrieval of nine atmospheric profile variables from a single set of collocated MW (AMSU or ATMS) and IR (AIRS or CrIS) measurements. We combine this approach with a rigorous error quantification and propagation to enable robust retrievals across a wide range of conditions of the global atmosphere. In Figure 17 we outline the CLIMCAPS sequential retrieval approach with a specific focus on O_3 at three steps. We discuss different aspects of the CLIMCAPS algorithm flow in Smith and Barnett (2019, 2020) respectively.

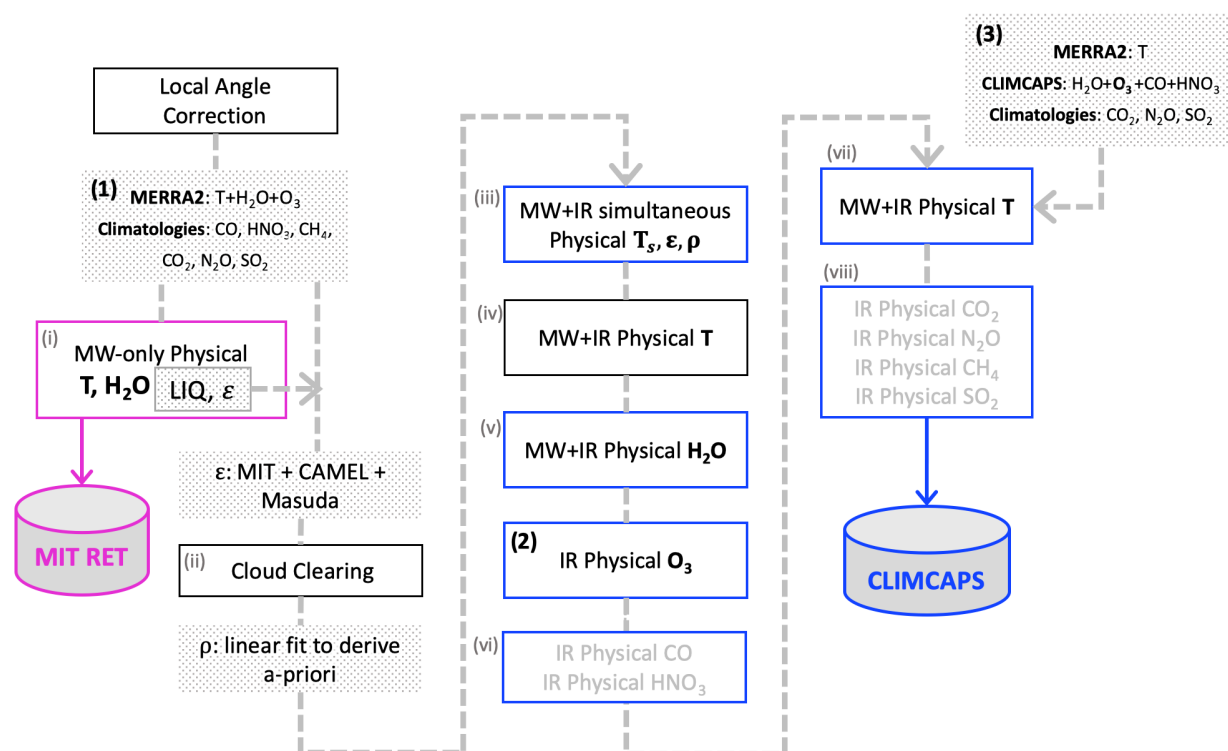


Figure 17: Flow diagram of CLIMCAPS sequential retrieval algorithm with a focus on O_3 along three steps as discussed in text below. The full algorithm flow diagram is available elsewhere and we discussed different aspects of algorithm flow in Figures 1 and 2 of (Smith and Barnett, 2019, 2020), respectively. Output from retrieval steps defines **air_temp** and the **mol_lay** subgroup in the netCDF product file. Output from the magenta step defines the **mw** subgroup.

Step (1) (Figure 17) initializes the nine profile retrieval variables – Temperature (T), H_2O , O_3 , CO, HNO_3 , CH_4 , CO_2 , N_2O and SO_2 – as their a-priori geophysical estimates and associated error covariance matrices. This set of a-priori estimates defines the background atmospheric state for (i) the MW-only retrieval step (Rosenkranz, 2001, 2006), (ii) cloud clearing, (iii) simultaneous surface parameter retrieval and, (iv) the first MW+IR temperature retrieval using the channel set indicated by grey dots in Figure 15b. From this step onward, the temperature a-priori is replaced with the output from step (iv) so that scene-dependent

knowledge of temperature and its associated uncertainties can correctly account for variation in the trace gas channel sets going forward. At (v), CLIMCAPS performs a joint MW+IR H₂O retrieval which similarly replaces the H₂O a-priori going forward. Now that temperature and H₂O are known, (2) O₃ is retrieved so that tropospheric temperature and H₂O interference in the O₃ channel set can be accounted for (Figure 15b). Then follows the retrieval of (vi) CO and HNO₃ in that order. At (vii) temperature is retrieved a second time using additional spectral channels (grey + black dots in Figure 15b) because scene-dependent knowledge of H₂O, O₃, CO and HNO₃ now enables a more accurate definition of the background atmospheric state and interference from these gas species in the T channel subset. Note that the a-priori and its associated error covariance matrix is exactly the same for the first and second T retrieval steps. The difference is that the H₂O, O₃, CO and HNO₃ a-priori profiles are replaced by the CLIMCAPS retrievals and the T measurement error covariance matrix is updated with estimates of scene-dependent uncertainty as described in (Smith and Barnett, 2019). The last step (viii) retrieves CO₂, N₂O, CH₄ and SO₂ in that order.

We wish to draw your attention to three areas where O₃ contributes to the CLIMCAPS system specifically. At (1) CLIMCAPS initializes MERRA2 as the O₃ a-priori with our empirical covariance matrix (Figure 22) as its error estimate. This definition of O₃ contributes to a number of steps as described above. At (2) CLIMCAPS retrieves O₃ using the ~1030cm⁻¹ channel subset (Figure 15b). For each retrieved O₃ profile, CLIMCAPS calculates a Bayesian a-posteriori error estimate that updates the a-priori error estimate with knowledge about scene-dependent uncertainty and instrument noise. At (3), the O₃ retrieval is now used as background state variable in the retrieval of temperature, while the O₃ a-posteriori error covariance matrix is added to the temperature measurement error covariance as described in Figure 1 of (Smith and Barnett, 2019). Not only does this second temperature retrieval benefit from scene-dependent knowledge of stratospheric O₃, it also benefits from a more accurate estimate of uncertainty about the O₃ state at the target scene.

5. O₃ information content and retrievals

As described in other chapters, we can calculate the total column information content for CLIMCAPS O₃ retrievals as the ‘degrees of freedom’ (DOF). Figure 18 depicts O₃ DOF from full spectral resolution CLIMCAPS-SNPP ascending orbits on 1 April 2016. The O₃ DOF from CLIMCAPS-Aqua and CLIMCAPS-NOAA20 are similar (not shown) and is in general around 2.0 for most of the globe. This means CLIMCAPS has sensitivity to O₃ at two distinct layers in the vertical atmosphere. When we plot the CLIMCAPS O₃ averaging kernels (Figure 19) we see that CLIMCAPS information content for O₃ has peaks in the stratosphere and upper troposphere.

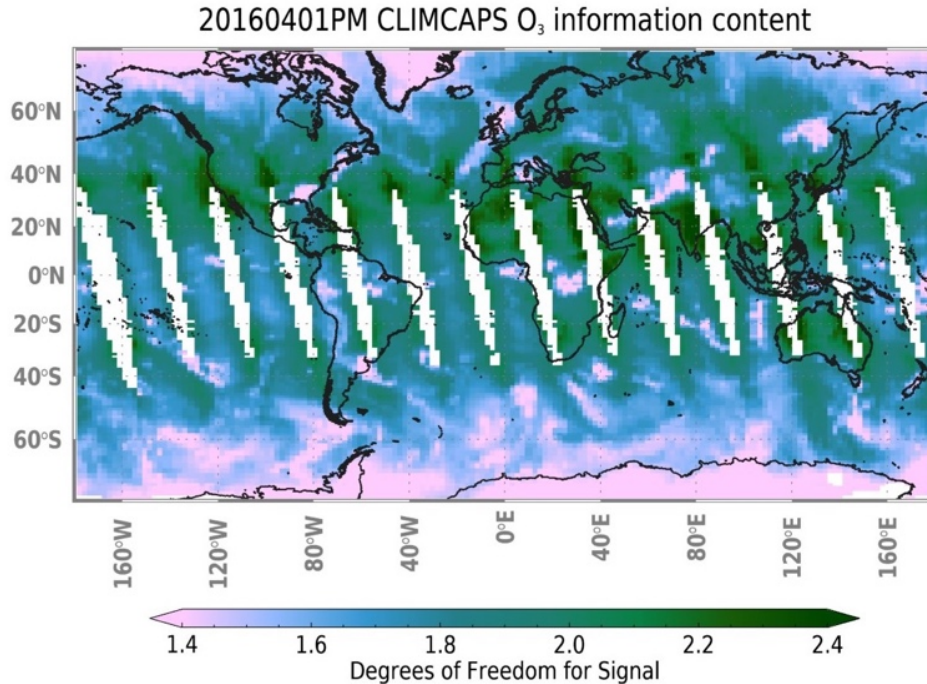


Figure 18: CLIMCAPS-SNPP degrees of freedom (DOF) for O₃ from full-spectral resolution CrIS at every retrieval scene from ascending orbits (13h30 local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about O₃ at every scene. For most of the globe, CLIMCAPS has O₃ $1.5 < \text{DOF} < 2.5$. We used the netCDF field **o3_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome.

Figure 19 depicts mean profiles (with standard deviation error bars) of (i) sensitivity to the true state given by the averaging kernel matrix diagonal, (ii) layer density retrievals and (iii) a-posteriori error for three latitudinal zones, (a) North Polar ($> 60^\circ\text{North}$), (b) Mid-Latitudes ($>30^\circ\text{N}$, $<60^\circ\text{N}$) and the Tropics ($>30^\circ\text{S}$, $<30^\circ\text{N}$). Note how there are fewer vertical error bars on the averaging kernel profile (blue line) compared to the retrieval (orange line) and error (yellow line) profiles. This is because CLIMCAPS averaging kernels are calculated on a reduced set of pressure layers, defined by a series of overlapping trapezoids.

The retrieval error profiles (yellow line) in Figures 19 represent the diagonal vector of the a-posteriori error covariance matrix (Figure 22) that CLIMCAPS generates for each retrieval variable at each field of regard. This error profile in the netCDF file (**mol_lay/o3_mol_lay_err**) has the same units [molec/cm^2] as the retrieval profile (**mol_lay/o3_mol_lay**) so we could easily calculate the error as a percentage by dividing the error by the retrieval, multiplied by 100. This error does *not* represent the accuracy, bias, or error with respect to the true state, but instead is a representation of how much CLIMCAPS adjusted the a-priori error estimate, given system uncertainty. For all other retrieved variables, the CLIMCAPS a-posteriori error is greater than the a-priori error (yellow versus grey profiles in righthand panels of Figure 19). What does this mean? It means that the a-priori error covariance (Figure 22) is probably too low, or under-represents the real uncertainty in O₃, and specifically in the stratosphere where we see the largest difference between these two error profiles. Alternatively, it means that we need to re-evaluate

our O₃ channel selection to increase the signal-to-noise in the retrieval. CLIMCAPS retrievals do benefit from this updated O₃ error covariance matrix as described above.

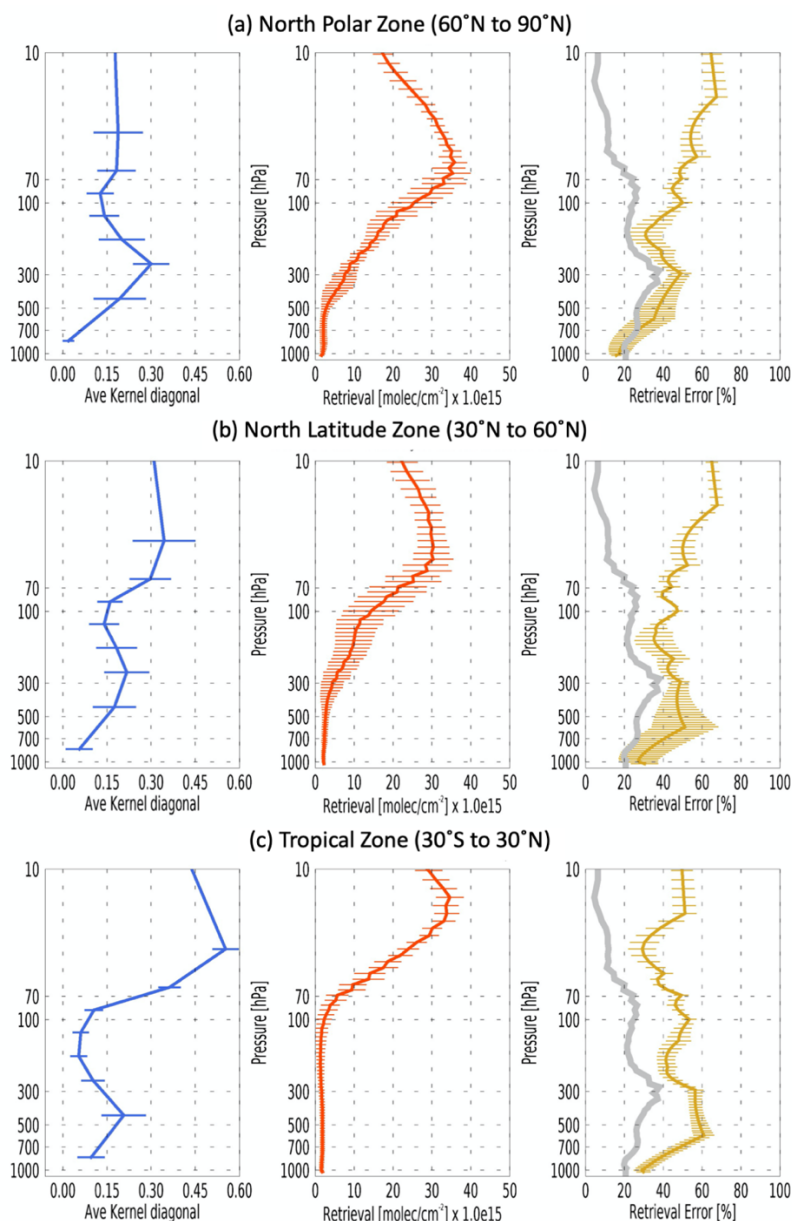


Figure 19: A diagnosis of CLIMCAPS-SNPP O₃ retrievals in (a) North polar zone [$>60^{\circ}\text{N}$], (b) North mid-latitude [$>30^{\circ}\text{N}$, $<60^{\circ}\text{N}$] and (c) Tropics [$>30^{\circ}\text{S}$, $<30^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile with error bars defined by the standard deviation at each pressure layer. [left column] CLIMCAPS O₃ averaging kernel matrix diagonal vector from netCDF field `ave_kern/o3_ave_kern` that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of O₃ in the atmosphere. [middle column] CLIMCAPS O₃ mean profile retrieval from netCDF field `mol_lay/o3_mol_lay` [molec/cm²]. [right column] CLIMCAPS retrieval error from netCDF field `mol_lay/o3_mol_lay_err` [molec/cm²] represented here as percentage $[\text{mol_lay/o3_mol_lay_err}]/[\text{mol_lay/o3_mol_lay}]*100$. CLIMCAPS uses an empirical a-priori error estimate, which is represented here by the thick grey line. In calculating these mean profiles, we filtered out all retrievals where `mol_lay/o3_mol_lay_qc(i,j) > 1`. We plot these profiles using the pressure layer array called `air_pres_lay`.

Users of the CLIMCAPS O₃ product can readily compare retrievals with their a-priori, MERRA2, since we include the space-time collocated MERRA2 fields in the product file. We give an example of how to diagnose CLIMCAPS O₃ retrievals with respect to their a-priori's in (Smith and Barnett, 2020).

Figure 20 and 21 are global maps of the CLIMCAPS O₃ retrieval fields gridded and averaged to 1.5-degree equal angle grids for the ascending orbits of CrIS/ATMS on SNPP on 1 April 2016. Total column O₃ [DU] (Figure 20) has contributions from stratospheric O₃ (Figure 21a) and tropospheric O₃ (Figure 21b).

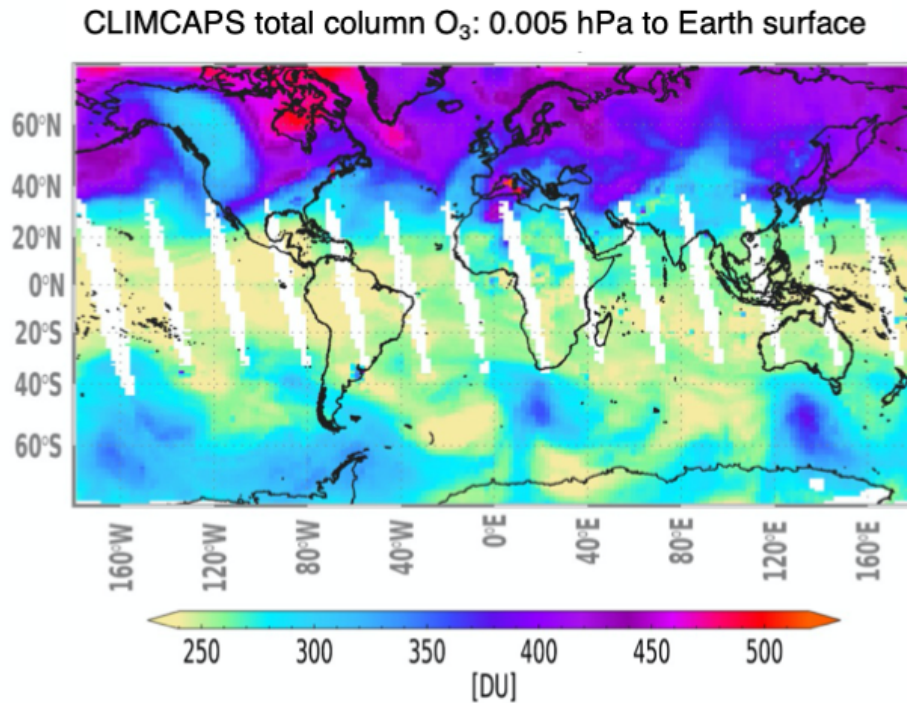


Figure 20: Global maps of daytime (ascending orbit; 13h30 local overpass time) CLIMCAPS-SNPP total column O₃ in Dobson Units [DU]. We generated this map using netCDF field **o3_tot** [kg/m²] multiplied by 4.67e+04 to convert to DU. We filtered out all values where **o3_tot_qc** > 1.

We made Figure 20 using the netCDF field **o3_tot** and filtered out all values where **o3_tot_qc** > 1. For the map in Figure 21a, we integrated all values in **mol_lay/o3_mol_lay** [molec/cm²] between vertical layers **air_pres_lay**(21)/100.0 = 10.26 hPa and **air_pres_lay**(39)/100.0 = 68.8 hPa. Figure 21b has **mol_lay/o3_mol_lay** [molec/cm²] integrated across **air_pres_lay**(56)/100.0 = 206.4 hPa and **air_pres_lay**(75)/100.0 = 487.2 hPa.

and quality control metrics in **mol_lay/o3_mol_lay_qc**. We integrated and averaged all values where the corresponding **mol_lay/o3_mol_lay_qc** > 1. Note that fewer retrievals are filtered out in the stratosphere (Figure 21a) due, predominantly, to the absence of clouds and H₂O.

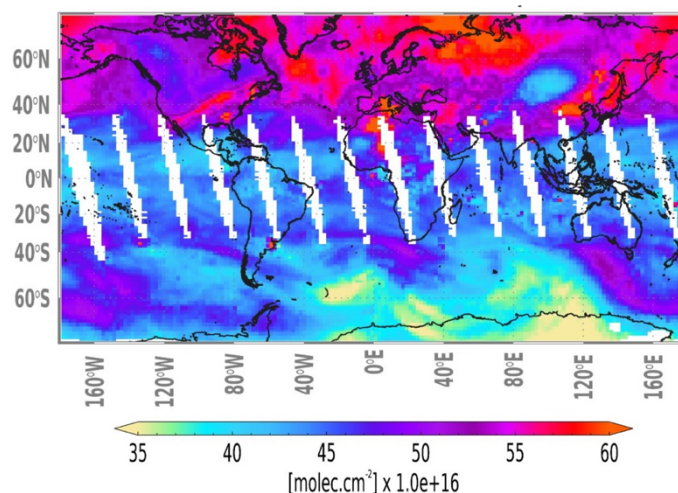
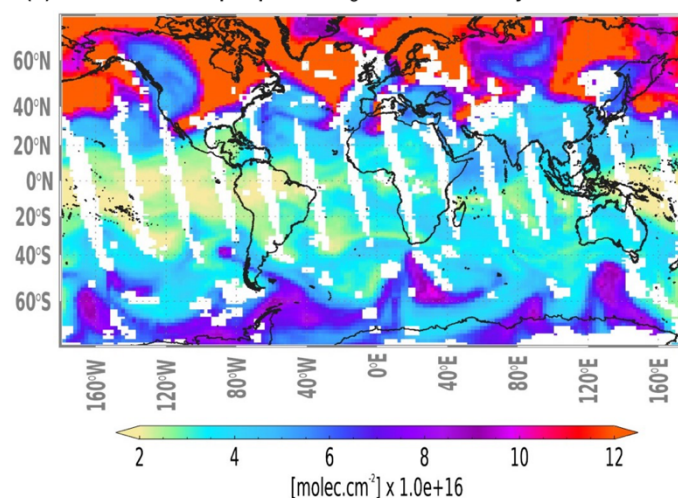
(a) CLIMCAPS stratospheric O₃ column density: 10.2–68.8 hPa(b) CLIMCAPS tropospheric O₃ column density: 206.4–487.2 hPa

Figure 21: Same as Figure 6 but for (a) stratospheric layer from 10.2–68.8 hPa, and (b) upper tropospheric layer from 206.4–487.2 hPa. For this plot, we used O₃ from the netCDF field **mol_lay/o3_mol_lay** on 100 pressure layers (**air_press_lay**) and summed all values for (a) across layers 4 to 39 and for (b) across layers 56 to 75. We removed all retrievals where **mol_lay/o3_mol_lay_qc** > 1.

6. Ozone a-priori

CLIMCAPS employs the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA2; Gelaro et al., 2017; GMAO, 2015; Wargan et al., 2017) as O₃ retrieval a-priori. CLIMCAPS converts MERRA2 from its native 72 pressure levels to column density [molec/cm²] on the standard 100 retrieval layers (**air_press_lay**). CLIMCAPS then interpolates these MERRA2 profiles spatially and temporally to match the cloud-cleared CrIS/AIRS measurements. We describe the benefits of employing MERRA2 as a-priori in Section 2.2.3 of Smith and Barnett (2019) and compare it against the a-priori approach taken in AIRS V5 and V6.

We derived the a-priori error covariance matrix $\delta O_3 \delta O_3^T$ (Figure 22) empirically from an ensemble of co-located MERRA2 and ECMWF O_3 profiles as percent difference: $(ECMWF - MERRA2)/ECMWF$. This ensemble was made up of 233,135 profiles from four global days in 2015, 1 January, 1 April, 1 July and 1 October. Before calculating the differences, we ran a 3-point running mean on each profile individually to smooth out fine structures. Then we filtered out all cases where the absolute value of the difference between ECMWF and MERRA2 was $> 50\%$ for H_2O and > 10 K for temperature. We applied no threshold to O_3 profile differences. We used the same ensemble to create a-priori error covariance matrices for CLIMCAPS temperature and H_2O retrievals. This is the only a-priori error estimate that we employ for CLIMCAPS O_3 . Each O_3 retrieval, no matter where on the globe or which day of the year, starts off with the error covariance matrix depicted in Figure 22 as the a-priori error estimate.

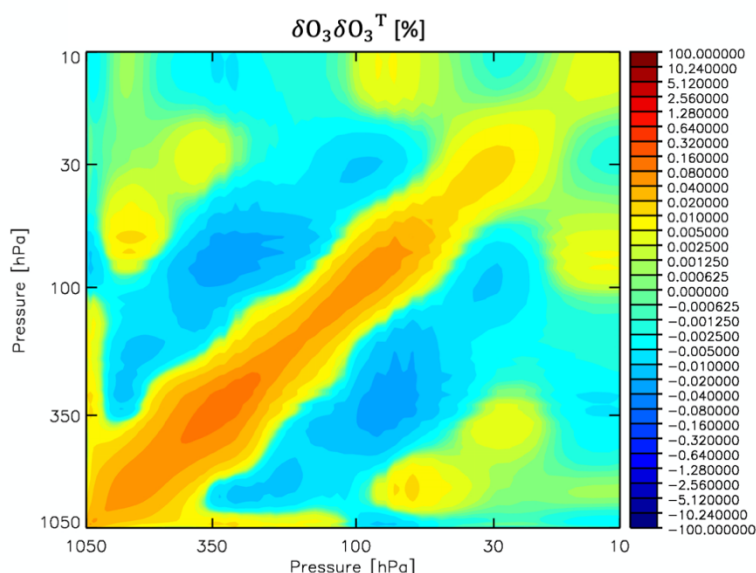


Figure 22: Empirical a-priori error covariance matrix used in CLIMCAPS V2 ozone retrievals. We derived this error matrix off-line from an ensemble of co-located ECMWF and MERRA2 profiles as the covariance, $\delta O_3 \delta O_3^T$, of $[ECMWF - MERRA2]/[ECMWF]$ to characterize the error as percentage (see section 2.2.4 of Smith and Barnett, 2019). The colorbar is in log-scale to enhance off-diagonal features.

We depicted the square root of the diagonal vector of this O_3 a-priori error covariance matrix (Figure 22) as the grey profile in Figure 19 and notice that it ranges between 20–40% in the troposphere and falls below 10% in the upper stratosphere. It is important to keep in mind that this error covariance matrix we derived does not represent the real error of MERRA2 O_3 , but merely an estimate given the assumption of ECMWF as the ‘truth’.

7. CLIMCAPS product field names relevant to O_3 applications

Within the netCDF files, the following fields are relevant for O_3 studies. Each CLIMCAPS file contains 45 scanlines along track (atrack) and 30 FOR along each scanline, or across track (xtrack). With O_3 profiles retrieved at each FOR on 100 pressure layers (air_pres_lay), the arrays have dimensions [atrack, xtrack, air_pres_lay].

Useful conversions:

$$1 \text{ kg/m}^2 = 4.4698\text{e}+04 \text{ DU} = 1.2547\text{e}+21 \text{ molec/cm}^2$$

$$1 \text{ DU} = 2.1414\text{e}-05 \text{ kg/m}^2 = 2.6868\text{e}+16 \text{ molec/cm}^2$$

$$1 \text{ molec/cm}^2 = 3.7219\text{e}-17 \text{ DU} = 7.9703\text{e}-22 \text{ kg/m}^2$$

– *Retrieved variables*

- **mol_lay/o3_mol_lay**: CLIMCAPS retrieval of O₃ layer column densities [molec/m²] on 100 pressure layers (**air_pres_lay**), where 1 molec/m² = 1.0e-04 molec/cm².

– *Derived variables*

- **o3_tot**: total column ozone as mass content [kg/m²], where 1 kg/m² = 4.670e4 DU.
- **o3_mmr**: mass mixing ratio, or the mass fraction of ozone in *dry* air [kg/kg] on 100 pressure levels (**air_pres**), where 1 kg/kg = 10e3 g/kg = 10e6 ppm.

– *Quality metrics*

- **ave_kern/o3_ave_kern**: O₃ averaging kernel matrix for every retrieval scene.
- **o3_dof**: The trace of the averaging kernel matrix as a measure of the number of independent pieces of information about O₃ at a target scene. For O₃, CLIMCAPS typically maintain a DOF below 2.5, which means that CLIMCAPS has sensitivity to the true state of O₃ in two vertical layers.
- ***_err**: The retrieval error estimate for column density (**o3_mol_lay**), total ozone (**o3_tot**) and mass mixing ratio (**o3_mmr**). The error has the same units as the original retrieved or derived variables.
- ***_qc**: The quality control flag for for column density (**o3_mol_lay**), total ozone (**o3_tot**) and mass mixing ratio (**o3_mmr**). This is the same quality flag used in the retrieved and derived variables.

– *A-priori variables*

- **aux/fg_o3_mol_lay**: MERRA2 a-priori for O₃ retrieval in the IR retrieval step. These are the native MERRA2 O₃ profiles converted to column densities [molec/m²] and interpolated spatially, temporally and vertically to match the CLIMCAPS retrieval footprints. This field is included in the netCDF file so that you can readily compare the retrieval with its a-priori. Multiply by 1.0e-04 to convert to the more common form of [molec/cm²].

Section 4: Carbon Monoxide

CLIMCAPS retrieves profiles of carbon monoxide (CO) column densities [molecules/cm²] on 100 pressure layers. These are the standard pressure layers used in modern-era radiative transfer models and that CLIMCAPS employs in all its trace gas retrievals. You can integrate column layer densities across any pressure range (e.g., 200–700 hPa) or for the total column (surface to top of atmosphere) simply by adding the retrieved layer densities together. CLIMCAPS retrieves

CO from a set of hyperspectral infrared channels selected from cloud cleared AIRS or CrIS radiance measurements in the 2100 cm^{-1} wavenumber (or $4.7\text{ }\mu\text{m}$ wavelength) range.

1. How can I access CLIMCAPS CO retrievals?

CLIMCAPS CO retrievals are distributed within the main product file (in netCDF format) that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://disc.gsfc.nasa.gov/>).

2. Which retrievals should I avoid?

Generally, we recommend avoid using CLIMCAPS CO retrievals in the boundary layer, tropopause and stratosphere. For atmospheric pressures greater than $\sim 700\text{ hPa}$ and less than $\sim 200\text{ hPa}$ CLIMCAPS has very low observing capability for CO (see Figures 3 and 4 for more details). We assign a quality control (QC) index for each retrieval that you can use to filter out ‘bad’ retrievals. By far the biggest reason CLIMCAPS retrievals fail is due to clouds covering entire the field of regard. Infrared (IR) radiation is strongly affected by clouds. A top of atmosphere IR radiance of an overcast scene contains no information about the atmospheric state below the scene. In version 2.0, we do not have quality controls set up for CO specifically, but apply the same quality flag to all retrievals. Avoid using CO retrievals from the *normal spectral resolution* CLIMCAPS-SNPP system, since CrIS spectral resolution in CO absorption region is too low to accurately detect CO.

3. How do I interpret CO retrievals in cloudy or smoky regions?

CLIMCAPS performs cloud clearing on all field of regards where clouds were detected to allow retrievals past the clouds. CO retrievals at the edges of clouds, where the scene is partly cloudy, are reliable representations. Usually CO retrievals are marked ‘bad’ when a smoke plume covers the entire field of regard and affects IR radiances as if they were opaque cloud fields.

4. Which type of applications are CLIMCAPS CO good for?

We recommend that you use CLIMCAPS CO products for long-range pollutant transport (Smith et al. 2020) and assimilation into chemistry models. We advise against using CLIMCAPS CO retrievals in air quality applications or urban-scale monitoring. Use CLIMCAPS CO for regional or global monitoring to flag regions for analysis with final scale measurements and products.

It is relatively easy to measure CO with IR instruments from space because CO has a distinct spectral signature in the 2100 cm^{-1} wavenumber range. CLIMCAPS selects a subset of these channels from AIRS on Aqua and CrIS on SNPP and NOAA-20 to retrieve CO after retrieval of clouds, temperature, moisture and ozone. IR spectral channels are highly correlated signals about multiple atmospheric state variables, so even those channels that are sensitive to CO specifically are also sensitive to other variables, albeit to a lesser degree. In CLIMCAPS, we attempt to quantify all sources of spectral interference and uncertainty due to instrumentation and knowledge about the atmospheric state to make CO retrievals as robust as possible (Smith and Barnett, 2019).

CLIMCAPS calculates an averaging kernel matrix for every CO retrieval and writes this to the product file. CLIMCAPS averaging kernels are unique to each retrieval variable at each field of

regard and can be used to evaluate the sensitivity of the retrieval system to the true state of the atmosphere, as ‘signal-to-noise’ or ‘information content’ metrics. The trace of the averaging kernel matrix is known as the ‘degrees of freedom for signal’ (DOF) and quantifies the number of independent pieces of information CLIMCAPS has about CO in the vertical atmospheric column at any given point in time and space. In **Figure 1** below we plot CLIMCAPS CO DOF for the ascending orbits (13h30 local overpass time) of SNPP on 1 April 2016 to illustrate its spatial variation (high values in Northern hemisphere, with very low values in Southern Polar region ($< 60^{\circ}\text{S}$). CLIMCAPS CO DOF varies daily and seasonally.

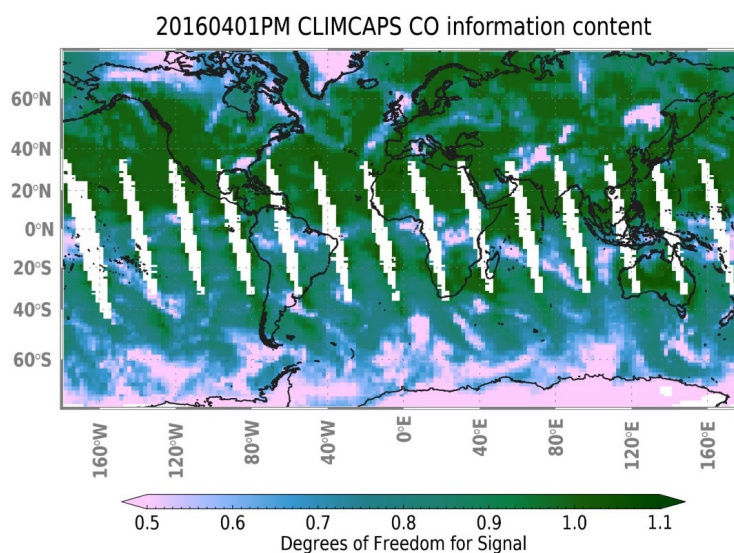


Figure 23: CLIMCAPS-SNPP degrees of freedom (DOF) for CO at every retrieval scene from ascending orbits (01:30 PM local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about CO at every scene. For most of the globe, CLIMCAPS has CO DOF of ~ 1 . We used the netCDF field `co_dof` and did not apply any quality filtering since DOF is not affected by retrieval outcome.

Note in **Figure 1** that CLIMCAPS maintains a fairly consistent pattern of CO DOF between 0.9–1.0 with values falling below 0.5 in the South Pole. A DOF ≈ 1.0 means that CLIMCAPS is sensitive to CO in a single vertical layer. A DOF ≈ 2.0 would indicate sensitivity to two vertical layers and so on. How should we interpret a DOF < 0.5 ? This means CLIMCAPS has very low observing capability of CO at that site and the retrieval contains mostly a-priori information and measurement uncertainty. In Figures 3 and 4 below we have a closer look at the vertical bounds of this single layer of CO sensitivity.

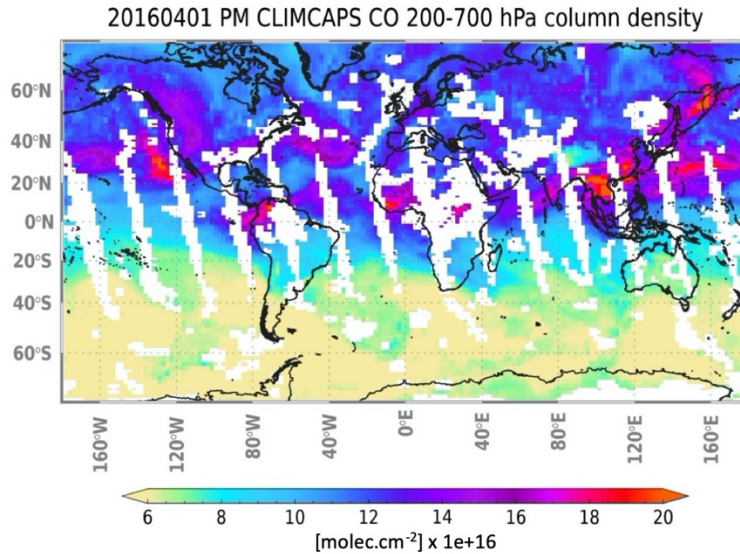


Figure 24: Global map of CLIMCAPS-SNPP retrieved CO tropospheric column density. CLIMCAPS retrieves CO on 100 pressure layers as column density, or number of molecules per cm^2 . Here we integrated all retrieved layers between 200–700 hPa to give an estimate of the mid-tropospheric CO load. We used the netCDF field `mol_lay/co_mol_lay`, integrated each profile into a tropospheric column density and filtered out all retrievals where their corresponding `aux/ispare_2` value was equal to one.

CLIMCAPS uses a static climatology as a-priori (also referred to as the first guess or background estimate) to retrieve CO from a subset of cloud-cleared IR radiance channels. This climatology consists of two profiles – one for the Northern and another for the Southern Hemisphere – to account for the strong latitudinal gradient in CO concentrations. CO also has seasonal variation due to the seasonality in biomass burning regimes from different parts of the world, which the CLIMCAPS climatology reflects by having two profiles (North vs South) for each month of a year, thus 24 in total (see [Section 2.2](#) for more details). We need to understand this to correctly interpret **Figure 2**. All the spatial variation in CO column density for this day, 1 April 2016, from West to East thus results from the IR measurements (information content and noise) and retrieval system design (e.g., uncertainty quantification and propagation) alone, not the background estimate used as a-priori.

In **Figures 3 and 4** we have a closer look at CLIMCAPS averaging kernels, CO profile retrievals and their associated errors. We plot these for CLIMCAPS retrievals from the full spectral resolution SNPP system on 1 April 2016. Figure 3 shows the mean profiles (with standard deviation error bars) for the North Polar region ($> 60^\circ\text{North}$), and Figure 4 for the Tropics ($30^\circ\text{South to } 30^\circ\text{North}$). We used the diagonal vector of the averaging kernel matrix as representation of the maximum sensitivity at each pressure layer. The blue line (Figures 3 and 4) is the mean of the diagonal vectors. Note how there are fewer vertical error bars on the blue line compared to the retrieval (orange line) and error (yellow line) profiles. This is because CLIMCAPS averaging kernels are calculated on a reduced set of pressure layers, defined by a series of overlapping trapezoids. The retrieval and its error is calculated on 100 pressure layers

because the radiative transfer model (SARTA) requires high vertical definition to calculate top of atmosphere IR radiances accurately.

Note how the averaging kernels show that CLIMCAPS has maximum sensitivity to CO broadly around 500 hPa. The closer an averaging kernel gets to zero, the more CLIMCAPS reproduces the a-priori without any information added by the measurements. Using a threshold of ~ 0.1 , we see that CLIMCAPS has sensitivity to CO within the vertical layer roughly from 200 hPa to 700 hPa. These are also the pressure bounds we used in calculating column density in **Figure 2**. The averaging kernel decreases sharply from 700 hPa to the Earth surface, which we interpret as CLIMCAPS having weak sensitivity to CO in the boundary layer. This means that CLIMCAPS does not have the ability to observe and monitor CO concentrations of interest to air quality forecasters, i.e., in the boundary layer air we all breathe every day. CLIMCAPS does, however, have the ability to observe CO concentrations in the mid-troposphere, and to do so over long distances day and night over the course of weeks, or for as long as CO concentrations are high enough to affect IR absorption lines at the top of atmosphere. We wrote about this capability for the CLIMCAPS sister system that runs in real-time at NOAA, namely NUCAPS, in an AMS paper (Smith et al. 2020).

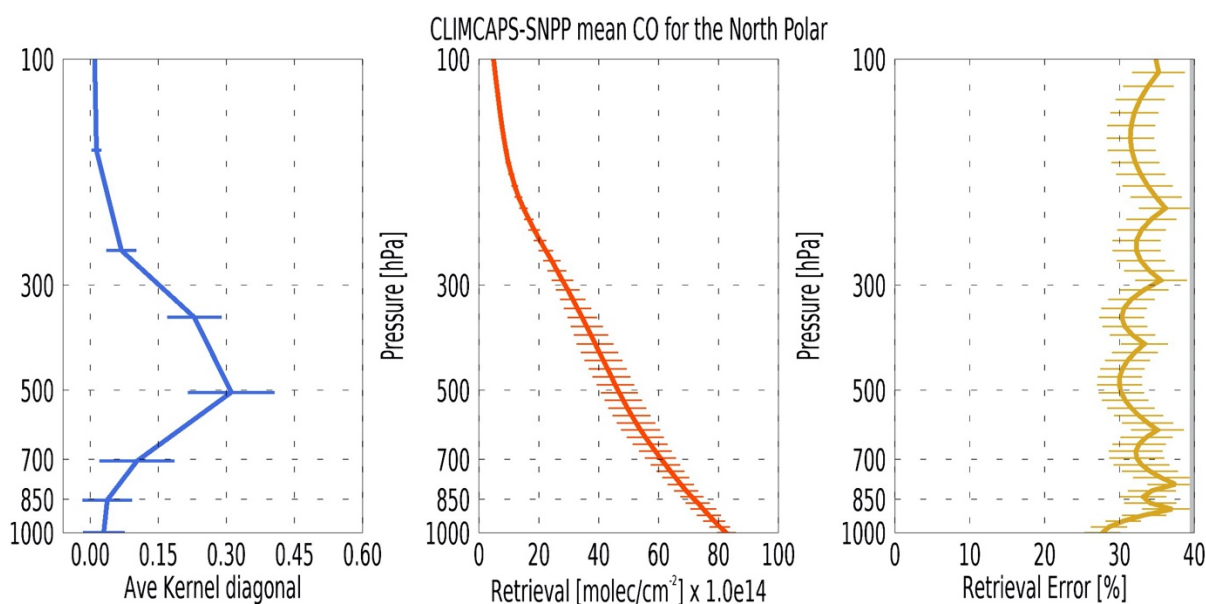


Figure 25: A diagnosis of CLIMCAPS-SNPP CO retrievals for the North Polar latitudinal zone [$>60^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile and error bars the standard deviation at each pressure layer. [left] CLIMCAPS CO averaging kernel matrix diagonal vector from netCDF field `ave_kern/co_ave_kern` that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of CO in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field `mol_lay/co_mol_lay` [molec/cm^2]. [right] CLIMCAPS retrieval error from netCDF field `mol_lay/co_mol_lay_err` [molec/cm^2] represented here as percentage $[\text{mol_lay/co_mol_lay_err}]/[\text{mol_lay/co_mol_lay}]*100$. CLIMCAPS uses a CO a-priori error of 40% as represented by the thick grey line. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals. In calculating these

mean profiles, we filtered out all retrievals where `aux/ispare_2(i,j) = 1`. We plot these profiles using the pressure layer array called `air_pres_lay`.

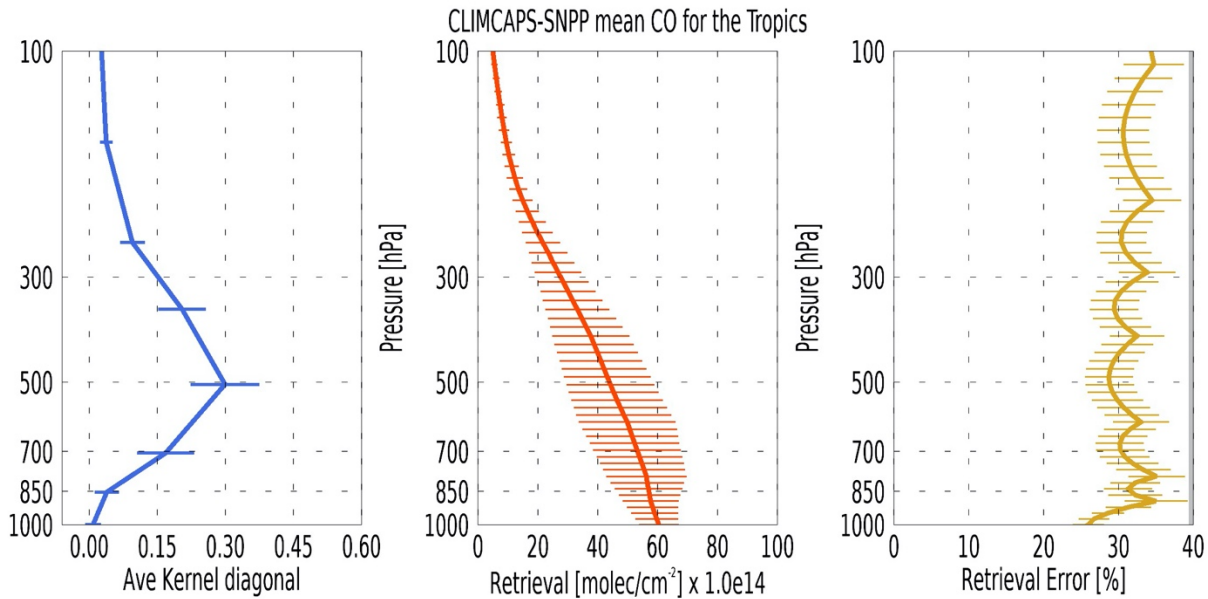


Figure 26: Same as Figure 3 but for the Tropical zone [30°S to 30°N].

Figure 26 shows that CLIMCAPS retrieves a higher variability in CO for the Tropics, which also corresponds to the stronger sources of CO in this latitude zone. The retrieval error we depict in the righthand panel (yellow line) of Figures 25 and 26 represents the diagonal vector of the a-posteriori error covariance matrix that CLIMCAPS generates for each retrieval variable at each field of regard. This error profile in the netCDF file (`mol_lay/co_mol_lay_err`) is in the same units [molec/cm^2] as the retrieval profile (`mol_lay/co_mol_lay`) so we could easily calculate the error as a percentage by dividing the error by the retrieval, multiplied by 100. This error does *not* represent the accuracy, bias, or error with respect to the true state, but instead is a representation of how much CLIMCAPS improved upon the a-priori error estimate.

CLIMCAPS defines a 40% error for its climatological a-priori, which we indicate by the thick grey line in the righthand panels of Figures 3 and 4. Here we see that CLIMCAPS reduces the a-priori error to ~30% within the pressure layers defined by the averaging kernel. You can use this error estimate as a sanity check to see whether the CLIMCAPS system reduced uncertainty in the climatological background knowledge of CO by adding information from IR measurements at that scene. If the a-posteriori error (also referred to as the ‘retrieval error’) exceeds the a-priori error (which is 40% at every retrieval footprint) at any given pressure layer or retrieval scene, then we can interpret it as the retrieval system being dominated by uncertainty for the target variable, given system design criteria such as channel selection, regularization parameters and uncertainty quantification. In Figure 26, we see this happening for some scenes in the Tropics at ~900 hPa (right panel, yellow error bar) where the information content also approaches zero (left panel, blue error bar). The a-posteriori error is a typical by-product of all systems using the Rodgers (2000) OE retrieval method.

5. Boundary layer adjustment

CLIMCAPS uses a standard 100-layer pressure grid to retrieve atmospheric variables from Earth surface (1100 hPa) to top of atmosphere (0.005 hPa). This pressure grid is required by radiative transfer models (SARTA for CLIMCAPS) to accurately calculate top of atmosphere hyperspectral IR radiances. CLIMCAPS uses the exact same pressure grid at every scene on Earth and accounts for surface pressure as a separate variable during radiative transfer calculations. The retrieved profiles are, however, reported on the 100-layer grid as a means to standardize the output. It is important that you adjust the bottom layer, i.e. that pressure layer intersecting the Earth surface as identified by `air_pres_nsurf` in the CLIMCAPS netCDF file, to accurately reflect the total number of CO molecules in the boundary layer.

You should do this boundary layer adjustment if you calculate total column densities or if you assimilate vertical profile retrievals. We describe the method for doing this adjustment in detail elsewhere.

This boundary layer adjustment is less relevant if you work with CO retrievals in the mid-troposphere, above the boundary layer such as depicted in Figure 24.

6. CO a-priori

The CLIMCAPS netCDF file does *not* have a field for the CO a-priori used in . CLIMCAPS uses a static climatological CO mixing ratio [ppb] a-priori that is interpolates latitudinally and for ‘day-of-the-month’ at run time. If your application depends on knowledge of the CO a-priori, e.g., data product assimilation and diagnostic comparisons, then you need to know how to calculate the CO a-priori for the retrieval scenes you are interested in.

CLIMCAPS CO a-priori is based on a 12-monthly climatology of two CO profiles per month from the Measurement of Pollution in the Troposphere (MOPITT; (Drummond and Mand, 1996)) instrument, which is onboard the Terra satellite. The two profiles represent the Northern Hemisphere (NH; $>15^{\circ}\text{N}$) and Southern Hemisphere (SH; $<15^{\circ}\text{S}$), respectively. In total there are, thus, 24 climatological CO profiles as baseline that capture seasonal variability in the two Hemispheres (Figure 27).

At runtime, and to avoid a sharp latitudinal break in the global representation of CO, CLIMCAPS performs a spatial interpolation between 15°N and 15°S (Figure 27) to introduce a gradual transition between North and South. This is necessary because there is a large difference in the mean background state between North and South in any given Month. When we plot these 12 climatological CO profiles for each hemisphere as a ‘curtain plot’ of month versus pressure [hPa] then we see the seasonal variation in each Hemisphere emerge (Figure 28). This seasonal pattern in CO concentrations are confirmed by in-situ measurements (Té et al., 2016). The NH experiences a maximum in tropospheric CO during March and April, and a minimum in October. The SH has a smaller amplitude in its seasonal cycle and experiences its peak CO concentrations in October and November.

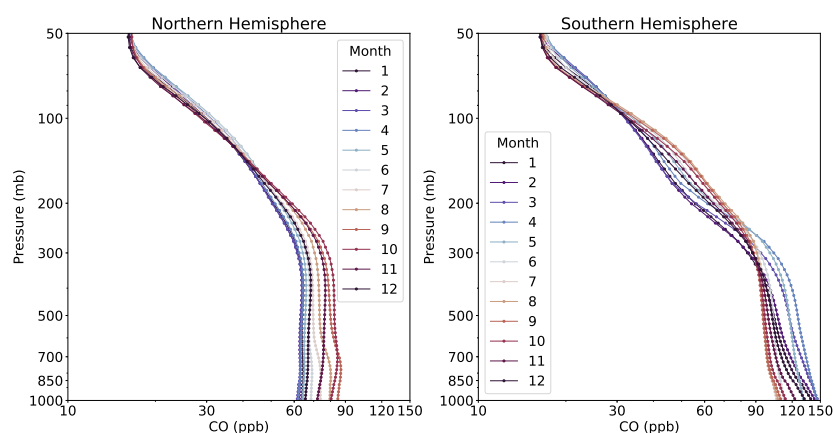


Figure 27: CLIMCAPS CO a-priori for the Northern and Southern Hemispheres for twelve months to capture seasonal variability in the background state. Month 1 is January, month 2 February and so on.

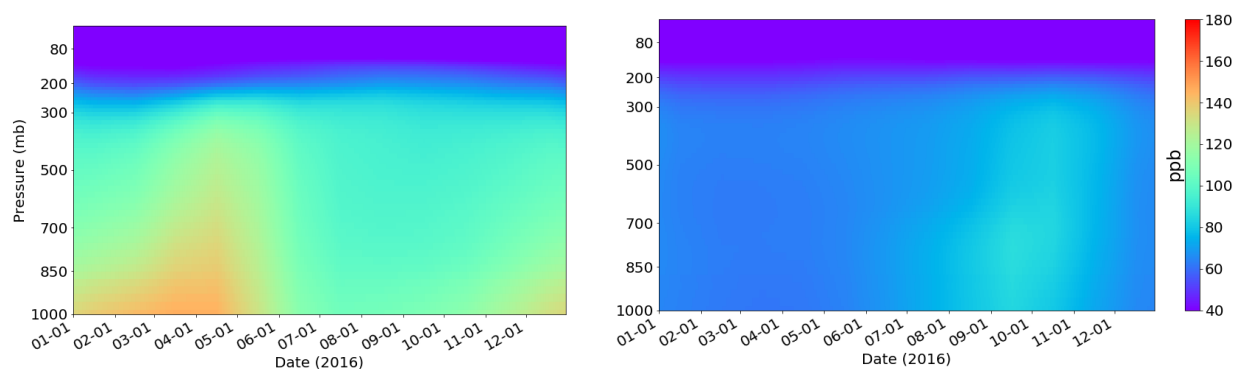


Figure 28: Seasonality of the profiles for the Northern Hemisphere (left) and the southern hemisphere (right) using the time interpolation scheme outlined in this section for the year 2016.

The NH and SH CO climatology profiles (Figure 28) are used “as is” in all retrieval scenes above 15°N or below 15°S, respectively. For retrieval scenes in the Tropics (15°S to 15°N), however, CLIMCAPS interpolates between the two climatological profiles to affect a smooth latitudinal transition using a weighted mean.

In **Box 1** below we describe how CLIMCAPS interpolates between the two hemispherical climatologies using ‘pseudocode’. If the retrieval scene latitude is below 15°S, then the NH weight ($Weight_{NH}$) is zero. If the retrieval scene latitude is above 15°N then the SH weight ($Weight_{SH}$) is zero. In between, there is 30° of separation, so the distance of this latitude from 15°N determines the weight. For instance, if the retrieval latitude is 7°S, then $Weight_{NH} = 0.27$ and $Weight_{SH} = 0.73$ according to the calculation depicted in the box below. Figure 30 shows what such an interpolation would look like for April (month 4) across all latitudes.

The CO climatologies are monthly averages of CO and CLIMCAPS sets the date of each to the middle day of the month (Figure 29). So, the profile for month 1 is given a date of January 16,

and month 2 as February 14. If the retrieval is not exactly on these dates, then CLIMCAPS calculates a weighted average between two monthly climatologies. For example, if the date is January 25, then CLIMCAPS interpolates between month 1 and 2 profiles; if the date is Jan 15, then CLIMCAPS interpolates between month 12 and 1.

We illustrate how to perform this temporal weighted average using pseudocode (**Boxes 2-4**). To simplify this calculation, convert the calendar dates of interest to Julian days so that you can calculate your dates as fractions of a year. For example, if the retrieval date is January 25, the date of the climatology profile ($\text{Time}_{\text{climatology}}$) for time_1 is January 16 (Julian day 14) and for time_2 is February 14 (Julian day 44). If it is a regular year with 365 days, the climatology time will be 0.041 and 0.126, respectively. The retrieval date, January 25 (Julian day 25), is a fraction of 0.068. The time weight ($\text{Weight}_{\text{time}}$) will then be $(25 - 16)/(44 - 16) = 0.310$ (**Box 2**).

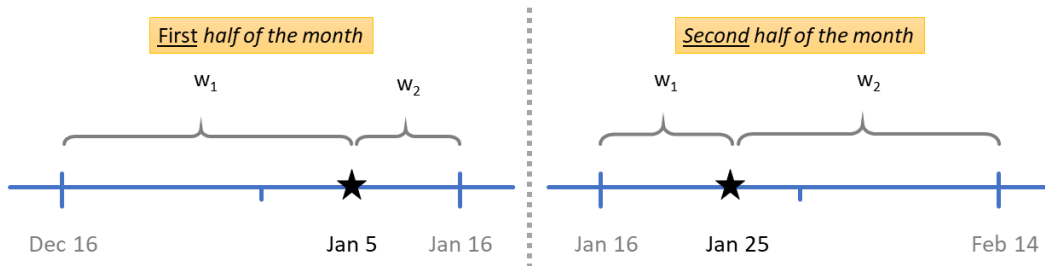


Figure 29: Sample time interpolation for CO first guess. The climatology dates are indicated by vertical lines, the star is the retrieval time, and w_i is the temporal weighting of each monthly climatology. If the retrieval is in the first half of month m , then the profiles for $m-1$ and m are used; if in the second half, then m and $m+1$ are used.

Box 1: First step in calculating a space-time interpolated climatology as CO a-priori at a target scene. Pseudocode for calculating a latitudinally weighted average of CLIMCAPS CO a-priori between 15°N and 15°S .

```

If ( $\text{latitude}_{\text{retrieval}} < -15^\circ$ ):
     $\text{Weight}_{\text{NH}} = 0.$ 
If ( $\text{latitude}_{\text{retrieval}} > 15^\circ$ ):
     $\text{Weight}_{\text{NH}} = 1.$ 
If ( $-15^\circ < \text{latitude}_{\text{retrieval}} < 15^\circ$ ):
     $\text{Weight}_{\text{NH}} = \text{Abs}(\text{latitude}_{\text{retrieval}} + 15^\circ)/30^\circ$ 
 $\text{Weight}_{\text{SH}} = (1 - \text{Weight}_{\text{NH}})$ 

```

Box 2: Second step in calculating a space-time interpolated climatology as CO a-priori at a target scene. Pseudocode for deriving the weight that will be used in time averaging in Box 4.

```

 $\text{Time}_{\text{climatology}}(\text{Month}) = \text{Julian day of year of the middle of the month}$ 
If  $\text{Time}_{\text{climatology}}(\text{Month}) > \text{middleOfMonth}$ :
     $\text{Time}_1 = \text{Month}, \text{Time}_2 = \text{Month}+1$ 
If  $\text{Time}_{\text{climatology}}(\text{Month}) < \text{middleOfMonth}$ :
     $\text{Time}_1 = \text{Month}-1, \text{Time}_2 = \text{Month}$ 

```

$$\text{Weight}_{\text{time}} = (\text{Time}_{\text{retrieval}} - \text{Time}_{\text{climatology}}(\text{Time}_1)) / (\text{Time}_{\text{climatology}}(\text{Time}_2) - \text{Time}_{\text{climatology}}(\text{Time}_1))$$

Now, you need to calculate the difference in monthly climatology according to **Box 3** for each hemisphere between time₂ and time₁.

Box 3: Third step in calculating a space-time interpolated climatology as CO a-priori at a target scene
Pseudocode for deriving differences between two standard climatology profiles based on the day of interest.

$$\Delta \text{Climatology}_{\text{month,SH}} = \text{Climatology}_{\text{CO}}(\text{Time}_2, \text{SH}, \text{Pressure}) - \text{Climatology}_{\text{CO}}(\text{Time}_1, \text{SH}, \text{Pressure})$$

$$\Delta \text{Climatology}_{\text{month,NH}} = \text{Climatology}_{\text{CO}}(\text{Time}_2, \text{NH}, \text{Pressure}) - \text{Climatology}_{\text{CO}}(\text{Time}_1, \text{NH}, \text{Pressure})$$

Lastly, combine the output from **Box 1** through **Box 3** according to the pseudocode equation in **Box 4**. The weighted time average of the climatology differences are added to the climatology for that month. The entire term is multiplied by the spatial weight for the hemisphere. The weighted northern and southern profiles are summed to get the first guess for CO, which is shown below.

Box 4: Fourth and final step in calculating a space-time interpolated climatology as CO a-priori at a target scene. Pseudocode

$$\text{Profile}_{\text{CO,SH}}(\text{Pressure}) = \text{Weight}_{\text{SH}} * (\text{Climatology}_{\text{CO}}(\text{Current Month}, \text{SH}, \text{Pressure}) + \text{weight}_{\text{time}} * \Delta \text{Climatology}_{\text{month,SH}})$$

$$\text{Profile}_{\text{CO,NH}}(\text{Pressure}) = \text{Weight}_{\text{NH}} * (\text{Climatology}_{\text{CO}}(\text{Current Month}, \text{SH}, \text{Pressure}) + \text{weight}_{\text{time}} * \Delta \text{Climatology}_{\text{month,NH}})$$

$$\text{Profile}_{\text{CO}}(\text{Pressure}) = \text{Profile}_{\text{CO,SH}}(\text{Pressure}) + \text{Profile}_{\text{CO,NH}}(\text{Pressure})$$

In Figure 30 we illustrate what a space-time interpolated CO a-priori would like for 1 April 2016.

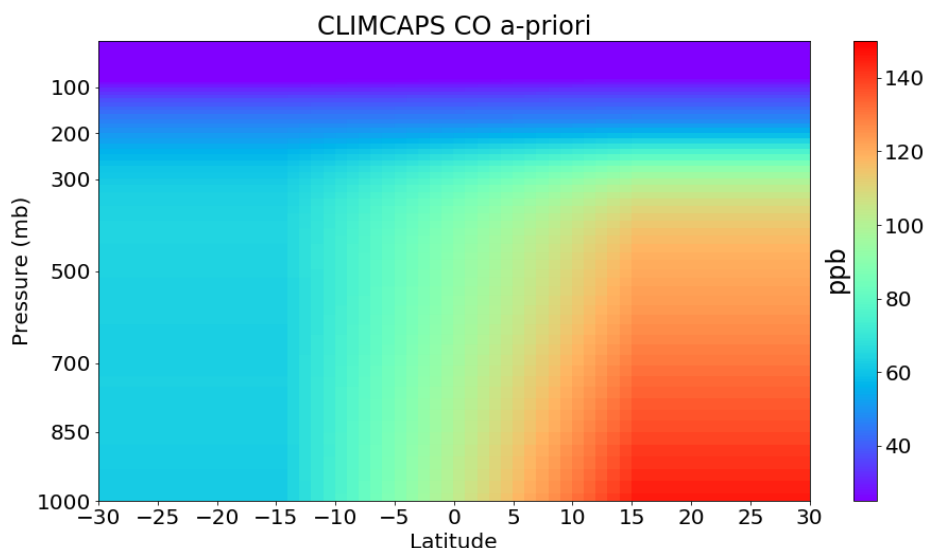


Figure 30: The CLIMCAPS CO a-priori over a fixed date (April 1, 2016) to demonstrate how latitude weighting varies by hemisphere. Poleward of 30°N and 30°S contain a constant profile, and a linear interpolation is performed between 30°S and 30°N.

If your application depends on column density CO retrievals [molec/cm^2] then you need to convert the CLIMCAPS CO a-priori from mixing ratio [ppb] to column density [molec/cm^2] using the method we describe here.

7. Convolution of a ‘truth’ profile to CLIMCAPS CO retrieval

If you want to assimilate CLIMCAPS CO retrievals then you would typically convolve your model fields to the CLIMCAPS vertical resolution using the averaging kernels and a-priori. The method for doing this is described in two papers (Maddy et al., 2009; Maddy and Barnett, 2008) and is the same method you would use if you compare radiosondes (or ozonesondes) to CLIMCAPS retrievals in validation studies. We describe this method in detail elsewhere.

8. CLIMCAPS product field names relevant to CO applications

Within the netCDF files, the following fields are relevant for CO studies. Each CLIMCAPS file contains 45 scanlines along track (atrack) and 30 FOR along each scanline, or across track (xtrack). With CO profiles retrieved at each FOR on 100 pressure layers (air_pres_layer), the arrays have dimensions [atrack, xtrack, air_pres_layer].

– Retrieved variables

- **mol_layer/co_mol_layer**: This is the column density CLIMCAPS profile CO retrieval [molec/m^2] on 100 pressure layers, **air_pres_layer** [Pa], from the top of the atmosphere (**air_pres_layer**(1) = 0.005 hPa) to Earth surface, which is either at the 100th layer (**air_pres_layer**(air_pres_layer)/100 = 1100 hPa) or at the scene-specific surface pressure (**air_pres_layer**(air_pres_layer_nsurf(*i,j*))). Note that CLIMCAPS is typically only sensitive to CO concentrations in a broad vertical layer peaking at 500 hPa, so you should evaluate this retrieval in combination with the averaging kernel (**ave_kern/co_ave_kern**) to understand how it differs from the a-priori. Multiply by 1.0e-04 to convert to the more common form of [molec/cm^2].
- **mol_layer/co_mol_layer_err co_mol_layer** This is the column density CLIMCAPS error estimate for the **co_mol_layer** retrieval [molec/cm^2]. This error estimate does not reflect the true bias or accuracy of the retrieval (it carries no knowledge of the ‘truth’) but instead gives an a-posteriori estimate of how well the Bayesian OE retrieval fared, given the a-priori estimate. This error estimate should typically be lower than the a-priori estimate, which is 40% for the CLIMCAPS CO a-priori at all scenes.
- **cld_top_pres**: Cloud top pressure retrieval [Pa] for the CLIMCAPS footprint (or field of regard). While not related to CO, cloud top pressure is a useful parameter for studying smoke transport. If the plume is in the lower troposphere, the cloud top pressure can be a useful tracer of *where in the vertical column* the smoke is. We discuss this in more detail in Smith et al. 2020.
- **cld_frac_tot**: The effective cloud fraction can be used with cloud top pressure to track cloud plumes. Even when not studying smoke, high cloud top fractions can also impact uncertainty in the retrieval, particularly when there is high horizontal variability in CO. Thus, cloud top fraction can also be useful for diagnosing CO retrieval uncertainty.

– *Derived variables*

- **co_mmr_midtrop**: Carbon monoxide mass mixing ratio to dry air [g/g] at 500 hPa. Use this field for a quick look of CO.
- **co_mmr_midtrop_qc**: The quality control flag for carbon monoxide mixing ratio. This is the same quality flag used in the mol_lay/co_mol_lay_qc retrieval.

– *Quality metrics*

- **mol_lay/co_mol_lay_qc**: 100-layer quality control flags with 0 = good, 1 = suspect, 2 = bad.
- **ave_kern/co_ave_kern**: CO averaging kernel matrix for every retrieval scene.
- **co_dof**: The trace of the averaging kernel matrix as a measure of the number of independent pieces of information about CO at a target scene. For CO, CLIMCAPS typically maintain a DOF below 1.5, which means that CLIMCAPS has sensitivity to the true state of CO at a target scene in one vertical layer.
- **aux/ispere_2**: a single quality flag per scene that we use as a quick yes/no flag.

Section 5: Carbon Dioxide

CLIMCAPS retrieves profiles of CO₂ mixing ratio [ppm] on the same fixed 100 vertical pressure layers and spatial footprints as all the other trace gas species from a few dozen cloud cleared AIRS or CrIS channels in the **long-wave infrared** (IR) band around 666–750 cm⁻¹ wavenumber (~14.3 μm).

1. How can I access CLIMCAPS CO₂ retrievals?

CLIMCAPS CO₂ retrievals are part of the main Level 2 product file that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://disc.gsfc.nasa.gov/>).

CLIMCAPS retrieves CO₂ as mixing ratio [ppm] on pressure layers (**air_pres_lay**) but the Level 2 product file reports CO₂ as volumetric mixing ratio (**co2_vmr**) on pressure levels (**air_pres**), which is the **mole of CO₂ per mole of dry air [mol/mol]**. There are CO₂ retrievals at every retrieval scene (~50 km at nadir and ~150 km at edge-of-scan), twice a day from each instrument ascending and descending orbits. Even though the product is distributed as Level 2, we strongly advise against using CLIMCAPS CO₂ the way it is available in these files, i.e., at 100 pressure layers for every retrieval footprint. Instead, we recommend that you integrate the profile vertically, and apply spatial and temporal averaging. Below we discuss how CLIMCAPS CO₂ retrievals at single footprint resolution are dominated by random variability in temperature, H₂O vapor, clouds, ozone and others. We illustrate how this random variability due to spectrally interfering variables can be reduced with data aggregation and highlight a few applications for which CLIMCAPS CO₂ may be useful once aggregated in space and time.

2. Can I use CLIMCAPS CO₂ retrievals for studying climate trends?

We strongly discourage you from deriving long-term trends of temperature or CO₂ with CLIMCAPS products. Why? Because the IR spectral channels sensitive to CO₂ emissions are the same ones also sensitive to small temperature variation. This means that the retrieval of long-term trends in temperature – on the order of ~0.1 K per decade – depends on accurate knowledge of CO₂ decadal patterns, which is difficult to know globally. In turn, the retrieval of CO₂ mixing ratio strongly depends on knowledge of temperature at every scene.

In CLIMCAPS V2 we take a different approach to AIRS V.7 by having a reanalysis model, MERRA2, as a-priori for temperature (Gelaro et al., 2017; GMAO, 2015). AIRS V7 uses a non-linear statistical regression (Milstein and Blackwell, 2016; Susskind et al., 2014) as temperature a-priori that is calculated at run-time, with the effect that the a-priori at each footprint is independent of its neighbors. There is, thus, no a-priori spatial structure. MERRA2 temperature, on the other hand, has strong spatial structure and meso-scale gradients of temperature in the troposphere and stratospheres. With a MERRA2 as temperature a-priori, the CLIMCAPS temperature retrieval thus has spatial, temporal and vertical structure built into it. We argue that with an accurate estimate of temperature as starting, CLIMCAPS should have an improved ability to separate CO₂ from temperature in the spectral channels.

Infrared instruments such as AIRS and CrIS are sensitive to atmospheric temperature at multiple pressure layers through CO₂ emissions (Maddy, 2007; Strow and DeSouza-Machado, 2020). Even though both CO₂ and temperature have distinct spectral signatures (Figures 31, 32, 34), they are very difficult to separate spectrally because the IR channels typically used for retrieving temperature are also highly sensitive to CO₂ emissions. Stated differently, the longwave IR channels that give us information about CO₂ mixing ratio in the stratosphere (~666 cm⁻¹) and mid-troposphere (~730 cm⁻¹) are also highly sensitive to small variations in temperature at a target scene. CO₂ is a well-mixed gas, both spatially and vertically, which means that its variation from scene-to-scene and day-to-day is low. In CLIMCAPS, as in many other IR retrieval systems, we wrestle with the problem of how to retrieve a slow changing gas species from spectral channels with interference from fast changing variables such as temperature. We discussed a similar problem with CH₄ and H₂O in the mid-wave IR band elsewhere.

Here, we give a brief overview of the CLIMCAPS CO₂ product to help guide interpretation and application. For an in-depth look, we refer the reader to work by Eric Maddy on the challenges and approaches to CO₂ retrievals from AIRS (Maddy, 2007b, 2007a, 2009, 2010; Maddy and Barnett, 2008). The figures we present here are simplifications of the analysis Eric Maddy performed, to illustrate why we urge users to treat CLIMCAPS CO₂ products with caution, not to diagnose and contrast AIRS or CrIS instrument capability or deep-dive into the theory of signal-inversion.

3. Sensitivity of the infrared spectrum to CO₂

AIRS and CrIS each have hundreds of IR channels with information about many atmospheric variables. Despite the relatively high spectral resolution of these instruments, their channels remain highly mixed signals where each channel is sensitive to multiple variables at any given point in space-time. We design retrieval systems to separate these spectral measurements into

distinct atmospheric variables, but have to contend with the difficulty of quantifying uncertainty, interference, signal and noise for each channel and each retrieval variable. It is nearly impossible to do this accurately for all variables under all conditions. CLIMCAPS retrieves many IR absorbing trace gas species for a number of reasons; (i) to support Earth system process studies with coincident observations of many atmospheric variable, (ii) improve CLIMCAPS retrieval of temperature and H₂O and, (iii) enable the calculation of outgoing longwave radiation, a climate variable that requires knowledge of the full atmospheric state.

As mentioned, the retrieval of CO₂ and temperature are closely linked because their spectral signatures overlap in both the shortwave [~2150–2550 cm⁻¹] and longwave [~650–1100 cm⁻¹] IR bands (Figures 31 and 32). In CLIMCAPS, we select subsets of IR channels for each retrieval variable separately. For CO₂, we select those that characterize the core of CO₂ absorption lines in the longwave band to effectively compute the derivative of the spectral area which is proportional to the amount of CO₂ in the atmosphere. For temperature we select those channels that are sensitive to the sides of the CO₂ absorption lines because they are the ones with narrow vertical kernel functions (Kaplan et al., 1977). Channels on the sides of absorption lines also tend to have less sensitivity to CO₂. Our philosophy in CLIMCAPS is to select as many channels as possible with narrow kernel functions to promote high vertical resolution in the retrieval of a target variable, but at the same time promote stability (high precision) under all types of conditions of any given global day.

In Figures 31 and 32 we plot the spectral signatures for five CLIMCAPS retrieval variables, namely CO₂, temperature, H₂O, O₃ and N₂O, in the long- and shortwave IR CrIS bands, respectively. We calculate these with the stand-alone AIRS radiative transfer model known as SARTA (Strow et al., 2003) for the CrIS full spectral resolution mode. With spectral ‘signatures’, we mean spectral fingerprints, –kernels, –weighting functions or –Jacobians. Here we calculate spectral kernel functions, K , as the absolute change in brightness temperature (BT) given a change in state variable, i.e., $\delta y / \delta x$, where y is the simulated TOA radiance converted to BT [K] at scene temperature, and x is the state variable:

$$K = |F(x) - F(x - p)| \quad 4$$

where F is SARTA and p the perturbation factor. In Figures 31 and 32, we perturb the gas species by 1%, thus $p = 0.01(x)$, and $T(p)$ by 0.1 K, thus $p = 0.1$. Kernels are typically 2-D matrices with dimensions [$m \times n$], where m is the number of spectral channels and n the number of pressure layers. Our brute force kernels in Figures 31 and 32 have $n = 1$ because we perturb the target variable along a single broad layer for the sake of simplicity and illustration. In retrieval systems, $1 < n \leq 100$ to maximize information content and vertical resolution, given system constraints.

Notice in Figure 31, how the temperature kernel (blue) overlaps the CO₂ kernel (red), which means that CrIS channels sensitive to CO₂ emissions, are also sensitive to variation in temperature. More specifically, the temperature and CO₂ spectral signatures in Figure 31 show us that the spectral signature for a ~4 ppm (1% of ~400 ppm) change in CO₂ along the full vertical column is roughly the same shape and magnitude as a 0.1 K change in mid-tropospheric

temperature. The fact is that CO_2 does *not* change by 4 ppm on a daily basis from scene to scene but instead takes years to change by that much. Temperature, on the other hand, can easily vary by several Kelvins at small spatial scales and across the span of several minutes to hours. For CLIMCAPS V2, we selected channels for CO_2 retrieval from the $666 - 750 \text{ cm}^{-1}$ wavenumber ($\sim 14.3 \text{ }\mu\text{m}$) range. Second to clouds, $T(p)$ is the strongest source of uncertainty in CO_2 retrievals. None of the other variables, such as stratospheric O_3 (green), boundary layer H_2O or mid-tropospheric N_2O cause significant spectral interference in the $\sim 14.3 \text{ }\mu\text{m}$ region. There are some differences in CO_2 channel selection between CLIMCAPS-Aqua, CLIMCAPS-SNPP, CLIMCAPS-NOAA20 that we have not homogenized yet, but adopted straight from their legacy systems, such as AIRS V6. We refer the reader to our section that highlights system differences (i.e., CLIMCAIS-AIRS vs -SNPP vs -NOAA20) and issues that may affect the continuity of CLIMCAPS retrieval products across these different instruments and satellite platforms.

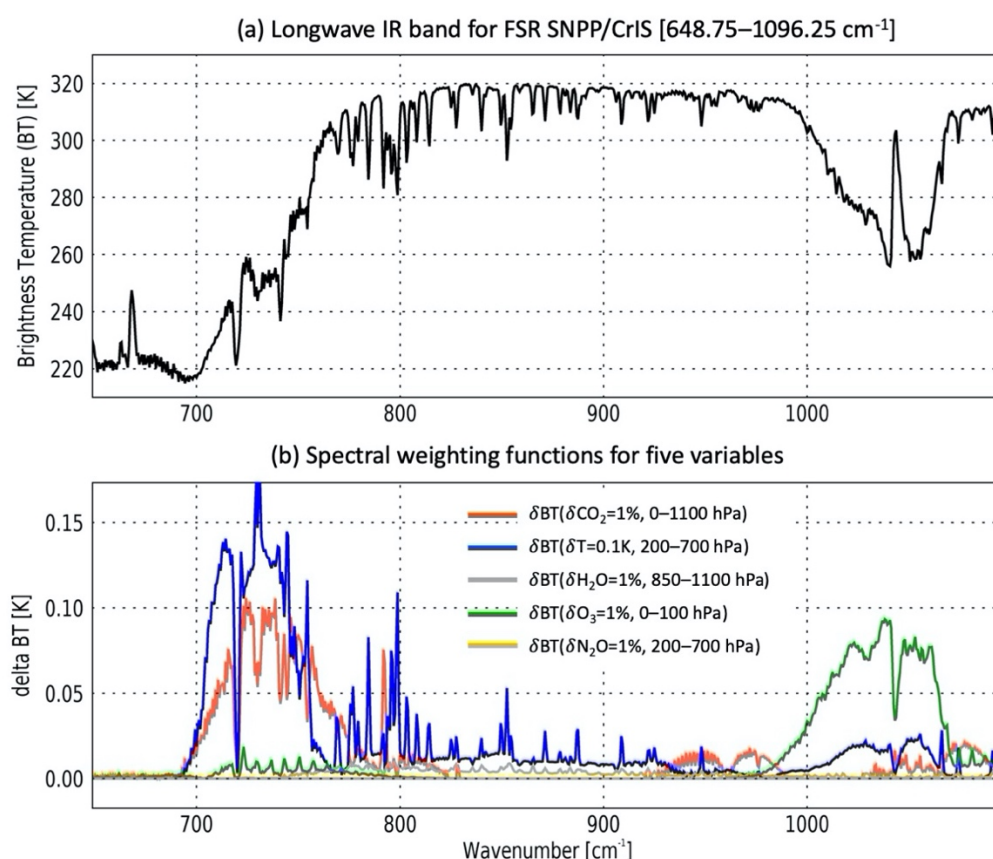


Figure 31: Simulated CrIS spectra using SARTA (Strow et al., 2003). (a) Longwave band [648.75 – 1096.25 cm⁻¹] of top of atmosphere CrIS spectrum at full spectral resolution (FSR), given CLIMCAP retrieval scene (1,1) from granule 104 on 1 April 2018. (b) Five spectral signatures or Kernel functions to illustrate channel sensitivity to a change in the target variable as $\delta\text{BT}/\text{dx}$ according to Eq. 1. They are, (red) 1% perturbation in column CO_2 [0 – 1100 hPa], (blue) 0.1 K perturbation in mid-tropospheric temperature [200 – 700 hPa], (grey) 1% perturbation for lower tropospheric H_2O [850 – 1100 hPa], (green) 1% perturbation to stratospheric O_3 [0 – 100 hPa], and (yellow) 1% perturbation to mid-tropospheric N_2O .

Even though CLIMCAPS V2 selects its CO₂ channels from the longwave band in the 14.3 μm range, we plot spectral signatures for the CrIS shortwave band in Figure 32 ($\sim 4.5 \mu\text{m}$) to illustrate what these spectral kernels (Eq. 1) look like for the same five variables. Note that CO₂ is not radiatively active in the mid-wave IR band, thus its omission in this chapter, but we refer the reader to our CH₄ chapter for a discussion of the mid-wave IR band.

Again, we see a strong spectral overlap between the temperature and CO₂ kernels, with a near-identical signature $\sim 2390 \text{ cm}^{-1}$ (Figure 32) for a daytime scene (ascending orbit, $\sim 13\text{h}30$ local overpass time). Unlike the longwave band (Figure 31), N₂O has a strong signature that overlaps with that of temperature and CO₂ in the $\sim 2200 \text{ cm}^{-1}$ range.

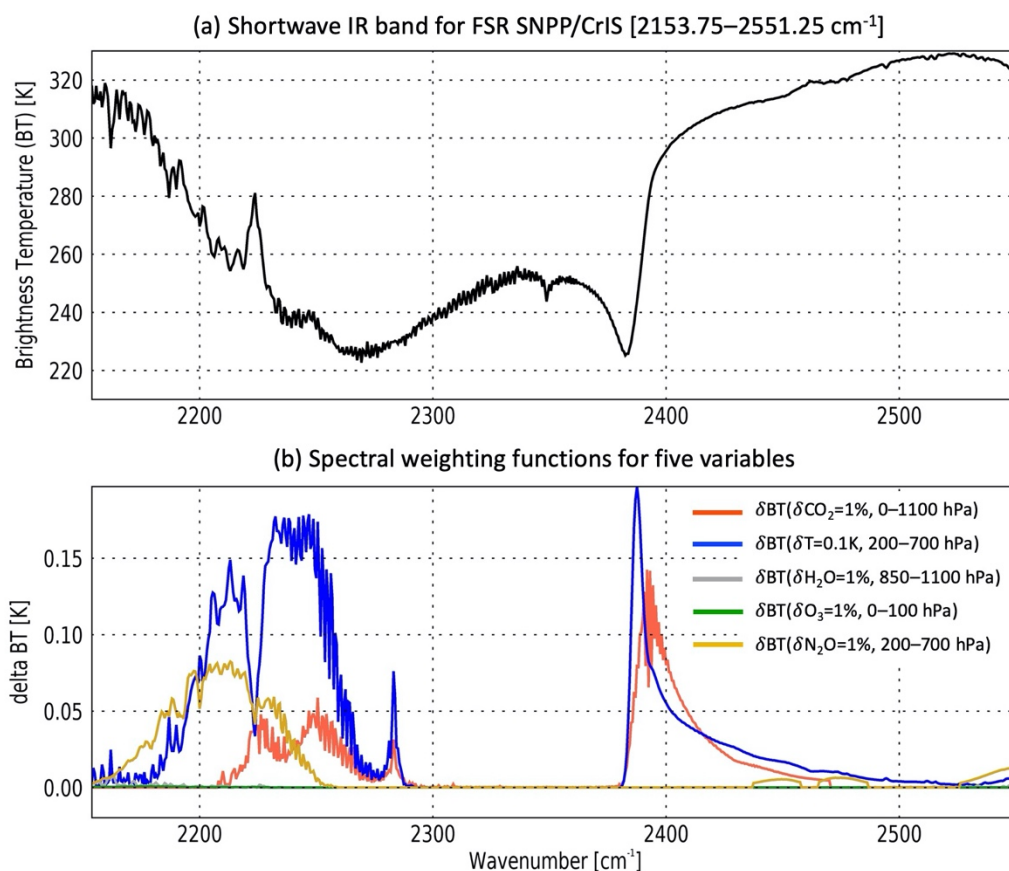


Figure 32: Same as Figure 31 but for the full spectral resolution (FSR) CrIS shortwave band [2153.75 – 2551.25 cm^{-1}]

We can visualize CLIMCAPS-SNPP information content as ‘degrees of freedom’ (DOF) for its CO₂ retrievals using the **co2_dof** field to further illustrate how temperature affects CO₂. In Figure 33a, with DOF on a 1.5-degree global grid, we see high spatial variability, much higher than what we can reasonably expect the $\sim 14.3 \mu\text{m}$ channels to have for CO₂ with $< 1 \text{ ppm}$ ($< 0.25\%$) variation across latitudinal zones on any given day. When we aggregate DOF (**co2_dof**) onto a 12-degree global grid (i.e., averaging all values binned into single grid cell), we gain a more realistic representation. Statistical averaging reduces random noise, so by

averaging CLIMCAPS CO₂ DOF, we reduce random uncertainty caused by channel sensitivity to large fluctuations in temperature.

Degrees of freedom quantifies the number of independent pieces of information that exists within a retrieval system about a target variable. In Figure 33, we see CO₂ DOF approximate 1.0 in the low latitudes, with values below 0.4 at high latitudes. This means that CLIMCAPS-SNPP V2 has at most one piece of information about CO₂ in the atmospheric column. This can be misleading because it does not mean that all the information is concentrated in one single pressure layer. Fractions of the total DOF can be from different parts of the atmosphere. In Figures 31 and 32, we calculated the spectral kernel functions for a 1% change in total column CO₂. We repeat the same calculation (Eq. 4) for the same scene in Figure 34, but this time perturb the CO₂ profile at three vertical layers to illustrate spectral sensitivity to CO₂ emissions from different parts of the atmosphere. In Figures 34a and 34b, respectively, we compare CO₂ spectral functions in the longwave [648.75–1096.25 cm⁻¹] and shortwave [153.75–2551.25 cm⁻¹] IR bands for (red) 1% perturbation in stratospheric CO₂ [0–100 hPa], (blue) 1% perturbation in mid-tropospheric CO₂ [200–700 hPa] and (gray) 1% perturbation in lower-tropospheric CO₂ [700–1100 hPa].

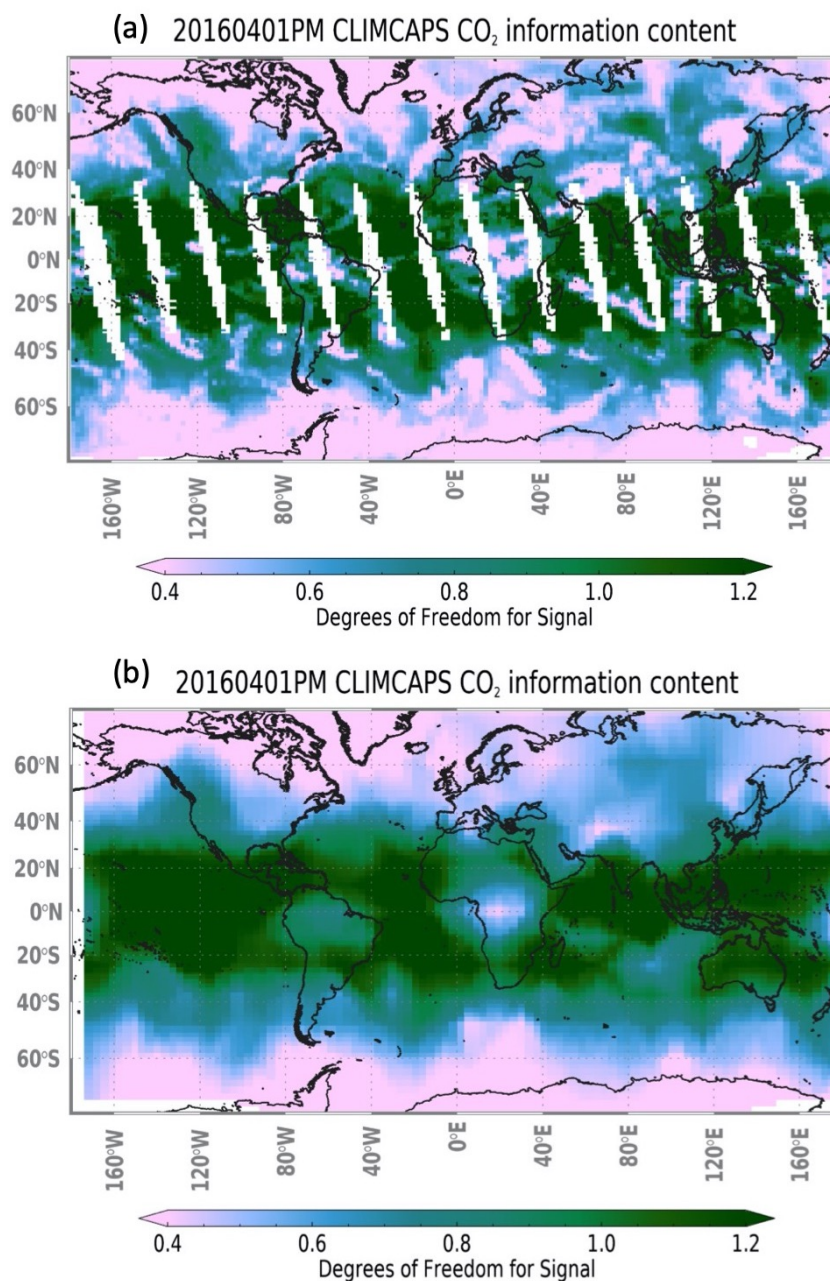


Figure 33: Information content as 'degrees of freedom' for CLIMCAPS-SNPP CO₂ retrievals (**co2_dof**) from full spectral resolution cloud cleared CrIS radiances for an ascending orbit (13h30 local overpass time) as a global equal-angle grid on (a) 1.5° resolution, close to single footprint size in the lower latitudes at edge of scan, and (b) 12° resolution as spatial aggregates (multiple values in the same grid cell were simply averaged). We did not apply any quality control filtering (**co2_vmr_qc**) since the averaging kernels (**ave_kern/co2_ave_kern**) from which DOF is derived are unaffected by the quality of the retrieval. DOF, instead, characterizes the potential a sounding system has in retrieving a target variable (Smith and Barnett, 2020). Results presented here are for CLIMCAPS-SNPP CO₂ retrievals on 1 April 2016.

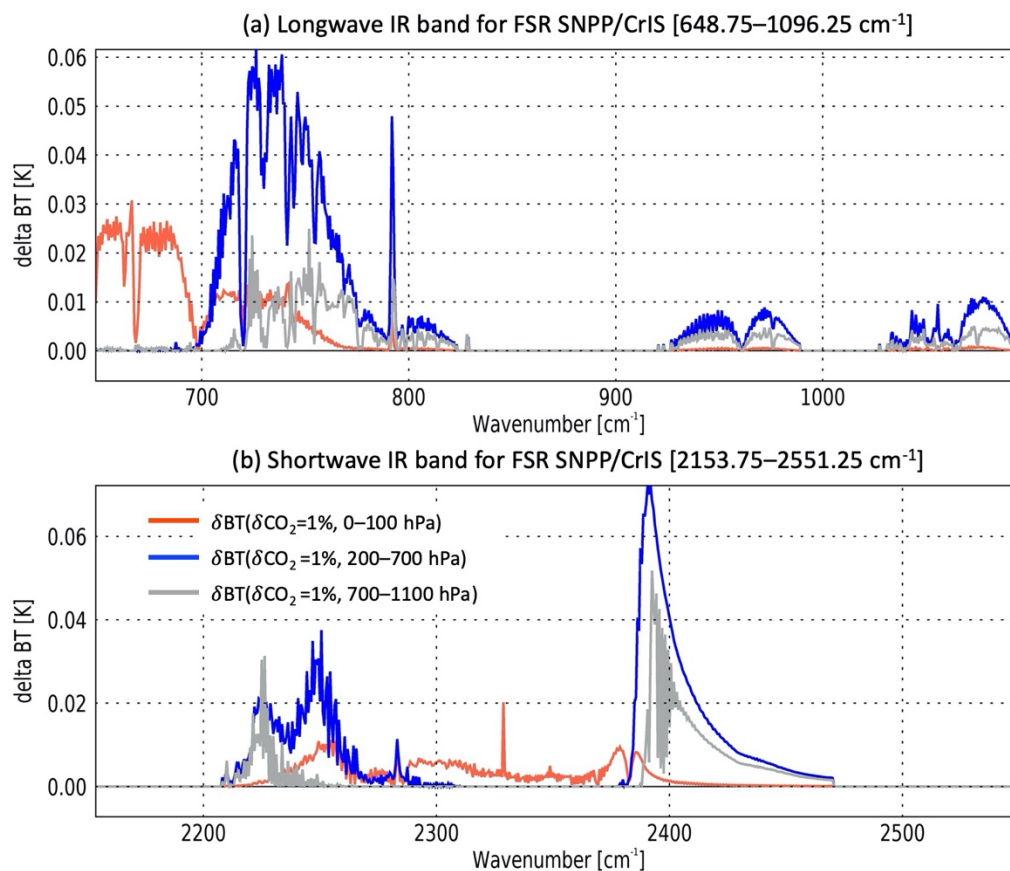


Figure 34: Same as Figure 31, but for (red) 1% perturbation in stratospheric CO₂ [0–100 hPa], (blue) 1% perturbation in mid-tropospheric CO₂ [200–700 hPa] and (gray) 1% perturbation in lower-tropospheric CO₂ [700–1100 hPa].

Figure 34a depicts a clear separation in kernels for CO₂ emissions in the stratosphere (red) versus mid-troposphere (blue). Sensitivity to lower tropospheric CO₂ emissions (grey) is mostly masked by the stronger sensitivity to CO₂ in the mid-troposphere (blue). So, while CLIMCAPS-SNPP has DOF ≈ 1.0 (Figure 33), the information we can retrieve about CO₂ emissions is typically smaller than that, closer to CrIS instrument noise.

CLIMCAPS uses ~ 57 CrIS channels selected from the longwave band in wavenumbers ranging 660–790 cm⁻¹ to help improve signal-to-noise, but weigh these channels inversely to their estimated random instrument noise (NEdT). In this sense, CLIMCAPS uses the systematic signal of the entire band to retrieve CO₂ and accounts for random instrument and retrieval noise.

Note that we are attempting to retrieve CO₂ with IR channels only, while retrieving temperature using IR and microwave (~ 57 GHz) channels. This has been the standard approach in AIRS Science Team in NUCAPS and AIRS V6. In addition to these measurements, CLIMCAPS does use MERRA2 as a-priori for temperature retrievals, and with that draws on the wealth of information assimilated from different sources, such as all the available space-based microwave and IR instruments, GPS-RO, radiosondes and model dynamics (i.e., the solution that balances

momentum, energy, and density conservation). With an accurate representation of temperature as background state variable in CO₂ retrievals (i.e., background state variables are used to calculate the kernel matrix, Eq. 4), we argue that CLIMCAPS may allow a better, more robust, separation of CO₂ and temperature from channels in the ~14.3 μm band because temperature is well known. When compared to CO₂ retrievals from systems that use statistical regressions as a-priori for temperature, such as AIRS V6, we already see a much improved, more realistic, spatial representation for CLIMCAPS CO₂ without the ‘salt-and-pepper’ effect of random noise that a statistical temperature a-priori introduces into subsequent CO₂ retrievals.

4. Regularization of the CO₂ retrieval

In this section we highlight some of the techniques we apply to regularize (or stabilize) the CO₂ retrieval. With this we wish to give the reader a sense of the CLIMCAPS algorithm and mathematical tools we apply to ensure a robust product. For a more thorough description of these techniques, we refer the reader to (Smith and Barnett, 2019, 2020)

CLIMCAPS applies a singular value decomposition (SVD) to its spectral kernel functions that are normalized by the measurement error covariance matrix at run-time to separate the signal (i.e., information about the target variable) from the noise (i.e., interfering species, which for CO₂ would be temperature). Each kernel decomposes into an eigenvector with associated eigenvalue, λ . The degree to which each eigenvector (or transformed kernel function) contributes to the retrieval is determined by λ , which we compare against a static threshold λ_c (Note: $\lambda_c \equiv 1/\text{co2wgt}$, with co2wgt defined as a constant variable in the CLIMCAPS namelists). When $\lambda/\lambda_c < 1$ the eigenvectors are damped proportionally as discussed in detail in (Smith and Barnett, 2020). When $\lambda/\lambda_c < 5\%$ CLIMCAPS simply returns the a-priori because the S/N is deemed too low. When $\lambda/\lambda_c > 1$ the eigenvectors are not damped and thus used without any regularization in the retrieval.

Most of this calculation is done dynamically at run-time for each radiance measurement and each retrieval variable, but the one static value we derive offline is co2wgt , which determines the critical threshold, λ_c , which in turn determines how much each eigenvector contributes to the retrieval and how much of the retrieval would ultimately depend on the a-priori. The smaller (larger) the co2wgt value, the more the retrieval depends on the radiance measurements (a-priori). Other factors affecting how CLIMCAPS weighs the contribution of the measurements to the final solution include the number of spectral channels selected for use in retrieval, how many trapezoid functions are defined for calculation of the K matrix and known errors like instrument noise, radiative transfer bias and so on.

In CLIMCAPS V2, $\text{co2wgt} = 0.38$ for AIRS and CrIS. We recognize that this may not be an optimal configuration since CrIS has apodized radiances with highly correlated S/N which may cause a higher degree of damping than AIRS radiance. We will address this in follow-on upgrades to the CLIMCAPS system to promote a stronger continuity in CO₂ retrievals between the AIRS and CrIS records. With $\text{co2wgt} = 0.38$, $\lambda_c = 1.6$, which means that only those eigenvectors with eigenvalues greater than 1.6 will contribute to the retrieval undamped. We adopted this value for co2wgt from the AIRS V6 system, but with MERRA2 as temperature a-priori in CLIMCAPS V2 and thus a better spatial structure as background, we need to reconsider

this value in future version as it may be too low, meaning that the information in each CrIS/AIRS measurement may be over-damped for CO₂.

Deriving an optimal value for `co2wgt`, however, is a classic Backup-Gilbert trade-off curve. For CLIMCAPS we want to find those values of `co2wgt` that minimize the error. If `co2wgt` is too large, the retrieval becomes the a-priori with minimal information contributed by the measurements. If `co2wgt` is too small, the retrieval depends primarily on the radiance measurements, as represented by the transformed kernel functions, or eigenvectors. But only those eigenvectors with $\lambda \geq 1.0$ contain information about the target variable. Eigenvectors with $\lambda < 1.0$ are dominated by noise. A `co2wgt` that is too small, allows these noisy eigenvectors to contribute to the retrieval undamped. If `co2wgt` = 0 it is equivalent to an unconstrained least square fit and the retrieval error can then exceed the error in the a-priori itself. In Table 4 we compare three different systems, all related to each other, but designed for different applications. There is NUCAPS (NOAA-Unique Combined Atmospheric Product System), based on the AIRS V5 code, that runs operationally at NOAA to support real-time weather forecasting (Esmaili et al., 2020; Weaver et al., 2019). Then, there is the legacy AIRS V6 system available at GES DISC for the full AIRS mission (AIRS Science Team/Joao Texeira, 2013).

Table 4: NUCAPS uses a linear regression for temperature T(p) retrievals, a global climatology for CO₂ that scales linearly over time, is available in near real-time (NRT), has no averaging kernels in its product file and no minor constituent detection flags. The AIRS V6 system uses a non-linear regression (or neural net NN) for T(p) retrievals, does not have a distinct CO₂ a-priori, is available in NRT with partial averaging kernels. CLIMCAPS uses MERRA2 as T(p) a-priori, has the same CO₂ climatology a-priori as NUCAPS, is not available in NRT but instead lags by a month (due to MERRA2), has the full averaging kernel matrices available at each retrieval footprint for each retrieval variable and has detection flags for minor constituent.

	NUCAPS	AIRS v.6	CLIMCAPS
T(p) a-priori	Regression	NN	MERRA2
CO ₂ a-priori	Global CO ₂ (t)	none	Global CO ₂ (t)
Latency	NRT	NRT	~1 month
Ave. Kernels	NO	Yes (partial)	Yes
Detection flags	NO	NO	Yes

5. Preparing CO₂ retrievals for applications

– *Boundary layer adjustment*

CLIMCAPS uses a standard 100-layer pressure grid to retrieve atmospheric variables from Earth surface (1100 hPa) to top of atmosphere (0.005 hPa). This pressure grid is required by radiative transfer models (SARTA for CLIMCAPS) to accurately calculate top of atmosphere hyperspectral IR radiances. CLIMCAPS uses the exact same pressure grid at every scene on Earth and accounts for surface pressure as a separate variable during radiative transfer calculations. The retrieved profiles are, however, reported on the 100-layer grid as a means to standardize the output. It is important that you adjust the bottom layer, i.e. that pressure layer intersecting the Earth surface as identified by `air_pres_lay_nsurf` in the CLIMCAPS netCDF

file, to accurately reflect the total number of CO₂ molecules in the boundary layer. We discuss how to do this in Chapter 3, Section 4.

– *Aggregation over pressure, space and time*

We calculate total column values for CLIMCAPS CO₂ mixing ratio profiles (**co2_vmr**) as the mean of **co2_vmr**($i, j, 1:100$) across all pressure layers, where i is the across-track (scanline) index and j the along-track (footprint) index (**Section 3**). The netCDF file reports **co2_vmr** in SI units [mol/mol], but we convert this to ‘parts per million’ with a multiplication by 10e+06 before averaging. In Figures 35 and 36, we depict column mean mixing ratio for CO₂ [ppm] across different space-time grids. Figure 35a has high spatial variability compared to Figure 35b, where mean(**co2_vmr**) values were calculated over larger grid cells; 12-degrees versus the 3-degree grid averaging in Figure 35a. Averaging reduces random variability, which explains why the spatial gradients in Figure 35b are smoother. Clouds are a major source random variability in CO₂ retrievals, thus the majority of spatial structure in Figure 35a can be attributed to scene-dependent uncertainty.

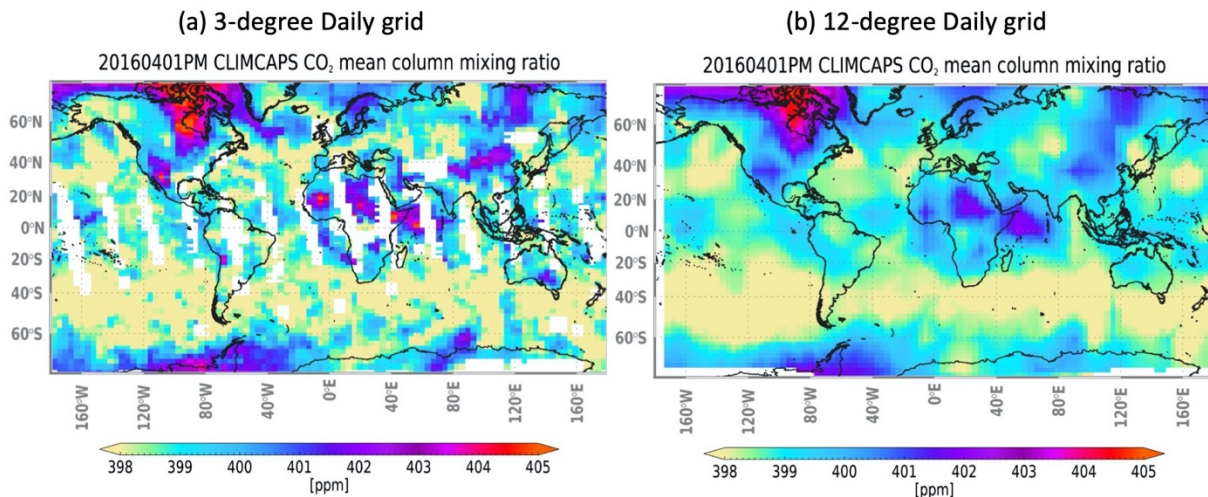


Figure 35: CLIMCAPS-SNPP CO₂ retrievals as column mean mixing ratio [ppm] for all ascending orbits (13h30 local overpass time) on 1 April 2016 aggregated to (a) 3-degree equal-angle grid and, (b) 12-degree equal angle grid.

In Figure 36, we calculated a monthly mean from the daily grids in Figure 35, by averaging both ascending and descending orbits for all days from 1–30 April 2016. We distinguish ‘spatial aggregation’ (averaging within grid cells) from ‘temporal aggregation’ (averaging across days) and note that temporal aggregation should be carefully considered. AIRS in Aqua and CrIS on SNPP and NOAA-20 only have two orbits per day – one at night and one during the day – so diurnal signals need to be considered. On any given day, these instruments have much more measurements about global spatial variability than temporal variability.

The true test of this product would be in its ability to discern CO₂ gradients, over time and space (Maddy, 2007b). More analysis is required to determine the skill of CLIMCAPS CO₂, whether it

matches gradients from other measurements over shorter timescales, and to better understand (inter/intra) product error correlations as a function of space-time (Maddy, 2007b).

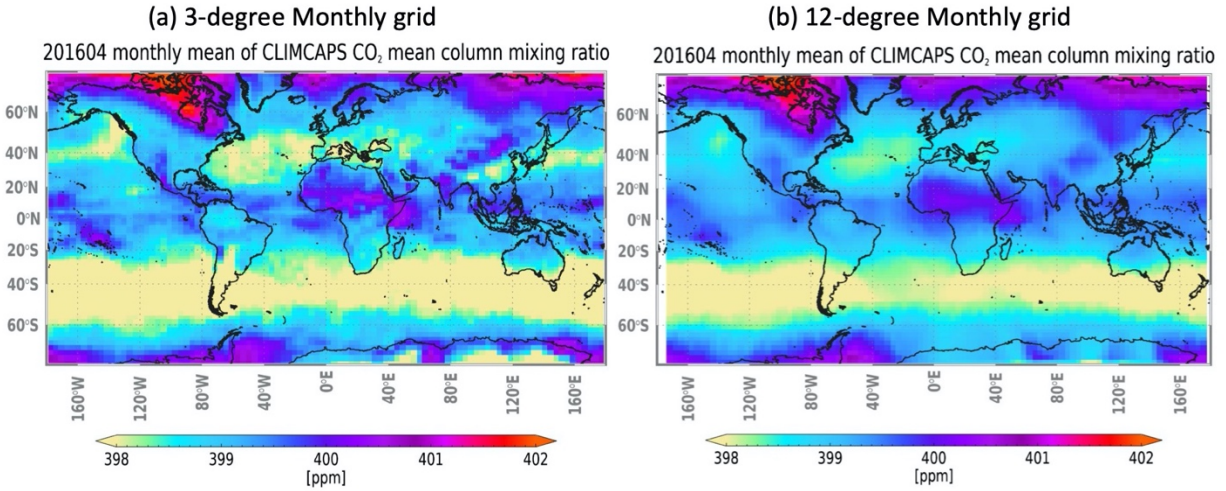


Figure 36: CLIMCAPS-SNPP CO₂ retrievals as column mean mixing ratio [ppm] averaged across all orbits (ascending and descending) from 1–30 April 2016 aggregated to (a) 3-degree equal-angle grid and, (b) 12-degree equal angle grid.

CLIMCAPS CO₂ is a new product and we will continue to evaluate and improve it. Here we demonstrate V2 capability. Our philosophy is to not hide any results, but to honestly communicate the strengths and weaknesses of our system, where there is room for improvement, and how observations should be interpreted. We encourage the reader to evaluate CLIMCAPS CO₂ products, communicate their results, not only to the science community in general, but share their findings, concerns and requirements with us as developers so that we may continue to improve observing capability.

– CO₂ a-priori

CO₂ is the only gas that CLIMCAPS retrieves as volume mixing ratio [ppm], not layer column density [molec/cm²]. The a-priori for CO₂ is a simple linear fit with no seasonal or latitudinal dependence, which is calculated as follows:

$$x_a = b + c(t_i - t_0) \quad 5$$

Where $b = 371.92429$, $c = 1.8406018$, $t_0 = 2002.0$, and $t_i = \text{year} + \text{month}/12.0$. Erik Maddy (Maddy, 2007a) developed this linear fit that we still use today. CLIMCAPS applies a 2% a-priori error.

There is room for improvement and we will consider two alternative approaches in future. One is being developed by Juying Warner at the University of Maryland. Warner (2020) argues for latitudinal and seasonal in a CO₂ climatology to improve the representation of gradients as a-priori for CLIMCAPS. Alternatively, on a given day we could create a CO₂ a-priori from spatially and temporally averaged CLIMCAPS retrievals from the previous day or two.

– *Applications CLIMCAPS CO₂ products are not suitable for*

Long term trends in CO₂ cannot be calculated from CLIMCAPS CO₂ retrievals. In essence, we will always have a maximum of about ~50% of the signal coming from our radiances and that is only in the easiest cases (clear, tropical lapse rates, etc.). We cannot reasonably expect to fully capture the trend (~2 ppm/year) or seasonal cycle (~3 ppm peak-to-peak) of CO₂ from IR measurements. We do, however, expect our CO₂ product to capture some of the inter-annual, seasonal, and regional variability.

When using CLIMCAPS CO₂ retrievals, keep in mind that they do not quantify boundary layer sources and sinks at city, county, or state level. Instead, they characterize CO₂ in the mid-troposphere at regional spatial scales as shown by the averaging kernels.

– *Consider using CLIMCAPS CO₂ products for these applications*

After spatial and temporal aggregation, we anticipate that CLIMCAPS CO₂ products can add value to a number of applications.

- (i) Data assimilation as discussed in (Engelen and Bauer, 2014).
- (ii) Continental-scale process studies: CLIMCAPS retrieves mid-tropospheric CO₂ at every cloud-cleared footprint with a global yield of 75% (i.e., retrievals that pass quality control) from ascending (day) and descending (night) orbits.
- (iii) Tracer-tracer correlations: CLIMCAPS retrieves O₃, CO, CO₂, CH₄, HNO₃, N₂O, and SO₂ from a single instrument measurement, which means that they are not only coincident in space and time but also thermodynamically consistent. Then, CLIMCAPS calculates minor constituent detection flags (isoprene, propylene, ethane, ammonia) to provide additional guidance on sources. See (Pan et al., 2006, 2007) for examples.
- (iv) Global quick looks: CLIMCAPS V2 generates atmospheric state variables from all measurements made by AIRS/AMSU on Aqua as well as CrIS/ATMS on S-NPP, and NOAA-20 (JPSS-1). It is a fast, robust system that can provide quick-look imagery that highlights areas for more in-depth study and targeted observations.
- (v) Verification of model transport models such as (Peters et al., 2007) and (Kawa, 2004).

– *CLIMCAPS product field names relevant to CO₂ applications*

Within the netCDF files, the following fields are relevant for CO₂ studies. Each CLIMCAPS file contains 45 scanlines along track (atrack) and 30 FOR along each scanline, or across track (xtrack). With CO₂ profiles retrieved at each FOR on 100 pressure layers (air_pres_lay), the arrays have dimensions [atrack, xtrack, air_pres_lay].

– *Retrieved variables*

- **aux/co2_vmr**: Volume mixing ratio in SI units [mol/mol] with respect to dry air. You can convert this to parts per million [ppm] by multiplying with 10e-06.
- **aux/clim_co2_mmr**: mass fraction of CO₂ [g/g] with respect to dry air.

– *Quality metrics*

- **aux/co2_vmr_qc**: profile quality control metrics ranging from 0 = good, 1 = suspect, 2 = bad.
- **ave_kern/co2_ave_kern**: CO₂ averaging

Section 6: Methane

CLIMCAPS retrieves profiles of CH₄ layer densities [molec/cm²] on the same fixed 100 vertical pressure layers and spatial footprints as all the other trace gas species. CLIMCAPS retrieves CH₄ from a subset of hyperspectral infrared channels selected from cloud cleared AIRS or CrIS radiance measurements in the fundamental ν_4 mid-wave infrared (IR) band around 1292-1306 cm⁻¹ wavenumber (7.6 μ m). We refer to this CLIMCAPS product where values are reported for each retrieval or scenes as ‘Level 2’ to distinguish it from the ‘Level 1’ radiance measurements and the ‘Level 3’ global, gridded aggregates.

1. How can I access CLIMCAPS CH₄ retrievals?

CLIMCAPS CH₄ retrievals are part of the main Level 2 product file that is generated and archived by the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://disc.gsfc.nasa.gov/>).

2. Which retrievals should I avoid?

Generally, avoid using CLIMCAPS CH₄ retrievals in the boundary layer, upper troposphere or stratosphere because CLIMCAPS has very low sensitivity to CH₄ at pressures greater than ~700 hPa. The CLIMCAPS product file has a range of quality control indices and diagnostic metrics. CLIMCAPS does not have quality control metrics specifically set up for CH₄, but instead adopts the logic that if retrievals of temperature and H₂O fail (usually due to uniform cloud fields), then CH₄ should also be considered a failed retrieval. CLIMCAPS retrieves CH₄ after it retrieves temperature, H₂O, O₃, HNO₃ and CO. You can customize the quality control filters according to different sources of uncertainty, such as those associated with the retrieval scene, instrumentation, background state or algorithm metrics. The most valuable algorithm metric to consider is the CH₄ degrees of freedom discussed below.

3. How do I interpret the spatial variability of daily CH₄ retrievals?

The 1300 cm⁻¹ wavenumber range has strong absorption features for both CH₄ and H₂O molecules. This means that the subset of IR channels CLIMCAPS use for CH₄ retrievals are also strongly sensitive to H₂O. The spectral signals of these two gases are thus highly correlated. We adopted a sequential retrieval approach in CLIMCAPS and select channels specifically sensitive to the target gas but even then, fail to completely remove the correlation between CH₄ and H₂O.

Atmospheric CH₄ has low variability from day to day, while H₂O has high variability even on an hourly basis. Much of the variability observed in CLIMCAPS CH₄ retrievals from orbit to orbit can (and should) be attributed to variation in H₂O. It is only when you aggregate CLIMCAPS CH₄ retrievals in space (4.0° grids) and time (monthly) that you average out random uncertainty from H₂O.

4. Which type of applications are CLIMCAPS CH₄ good for?

We recommend that you interpret Level 2 CLIMCAPS CH₄ products with caution since much of the daily variability can be attributed to H₂O variability despite efforts to minimize the correlation between CH₄ and H₂O in the inversion of spectral channels. The CLIMCAPS CH₄ product is thus not suitable for applications that require instantaneous, single-footprint observations. While, CH₄ has a lifetime of ~10 years and a long-term record of satellite observations could potentially be useful in studying global trends, we advise against employing CLIMCAPS CH₄ for that purpose because IR measurements in general have very low sensitivity to CH₄ in the boundary layer (where anthropogenic sources occur) and CLIMCAPS, as observing system, has low sensitivity to CH₄ in the mid- to upper troposphere due to a number of factors, such as choice in regularization parameters as well as error propagation. We envisage CLIMCAPS CH₄ retrievals to have value in atmospheric process studies where the full atmospheric state – CLIMCAPS retrievals of temperature, H₂O, trace gases and clouds – are considered.

Note that the main reason we retrieve CH₄ within CLIMCAPS is to improve the H₂O vapor retrieval in the troposphere because CH₄ absorption is significant in this spectral region. In version 2.0, CLIMCAPS uses the CH₄ climatology a-priori (Section 2.3) as background state variable in its H₂O retrieval; however, in future we may use the CLIMCAPS CH₄ retrieval to improve retrievals of tropospheric H₂O vapor.

IR instruments on space-borne platforms, such as AIRS on Aqua and CrIS on SNPP and NOAA-20, are not well suited to monitor CH₄ sources and sinks because their top of atmosphere measured radiances have weak sensitivity to atmospheric CH₄, which is limited to the mid- to upper troposphere (< 700 hPa). IR instruments have almost no sensitivity to CH₄ in the boundary layer. The difficulty in IR space-based CH₄ observation is compounded by the fact that the same spectral channels sensitive to CH₄ (~1300 cm⁻¹ wavenumber range) are also sensitive to H₂O vapor, a gas that exists in far greater quantities than CH₄ and results in a much stronger spectral signature (Figures 37 and 38).

In Figure 37a we plot the full spectral resolution CrIS mid-wave band with values simulated by SARTA using a CLIMCAPS-SNPP state retrieval – scanline 1, footprint 1, ascending granule 101 on 1 April 2018 – of temperature, H₂O vapor, trace gases, surface temperature and cloud parameters. In Figure 37b, we illustrate the spectral signatures of CH₄ and H₂O in this mid-wave infrared band as brightness temperature (BT) difference plots; the spectrum in Figure 37a is subtracted from a simulated spectrum using the same set of state variables, except for a percentage change in one of the target variables. In Figure 37 we investigate a 1% perturbation of CH₄ and H₂O vapor, respectively for the mid-troposphere (200–700 hPa) and lower troposphere (700–1100 hPa). For most spectral channels, H₂O is the dominant signal.

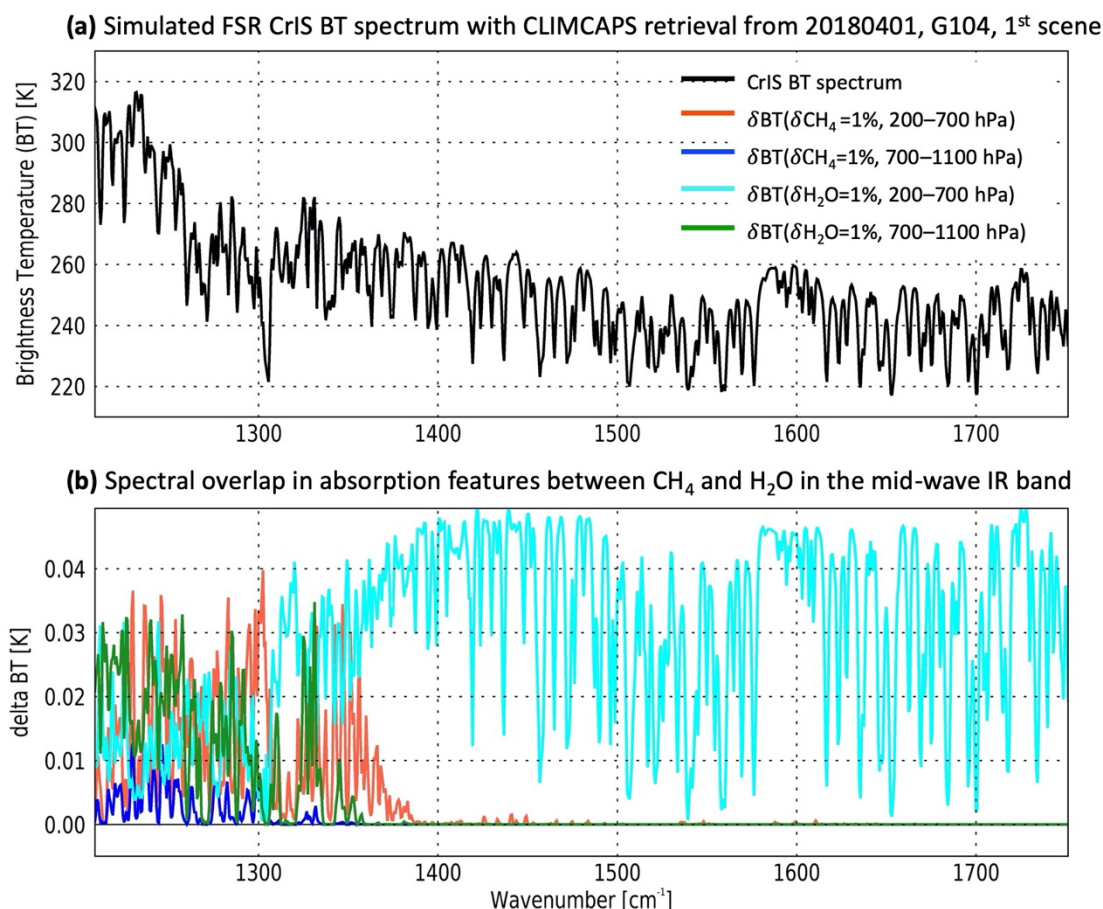


Figure 37: (a) SARTA simulated full spectral resolution CrIS Mid-Wave band (1209.75 – 1751.25 cm^{-1}) using a CLIMCAPS sounding retrieval as atmospheric state; specifically, the first retrieval (scanline=1, footprint=1) of granule number 104 on 1 April 2018 on an ascending orbit (13h30 local overpass time). The CrIS spectrum was simulated in radiance units [$\text{mW}/\text{m}^2/\text{steradian}/\text{cm}^{-1}$] and converted to brightness temperature [K] using scene temperature. (b) Absolute values of Brightness Temperature (BT) difference to illustrate absorption features for methane (CH_4) and water (H_2O) vapor given a 1% perturbation across different parts of the atmosphere. Red: 1% CH_4 perturbation for mid-troposphere, 200–700h hPa. Blue: 1% CH_4 perturbation for lower troposphere, 700–1100 hPa. Cyan: 1% H_2O perturbation in mid-troposphere, 200–700 hPa. Green: 1% H_2O perturbation in lower troposphere, 700–1100 hPa.

In Figure 38, we illustrate the strong spectral response in CrIS mid-wave IR channels when H_2O vapor varies from 1% to 10% (a range easily achieved on a daily basis) in the tropospheric column 200–1100 hPa. The red line in Figure 38, is the CH_4 spectral signature for a 1% perturbation along the same pressure range as H_2O . The CH_4 signal is spectrally overwhelmed by variation in H_2O .

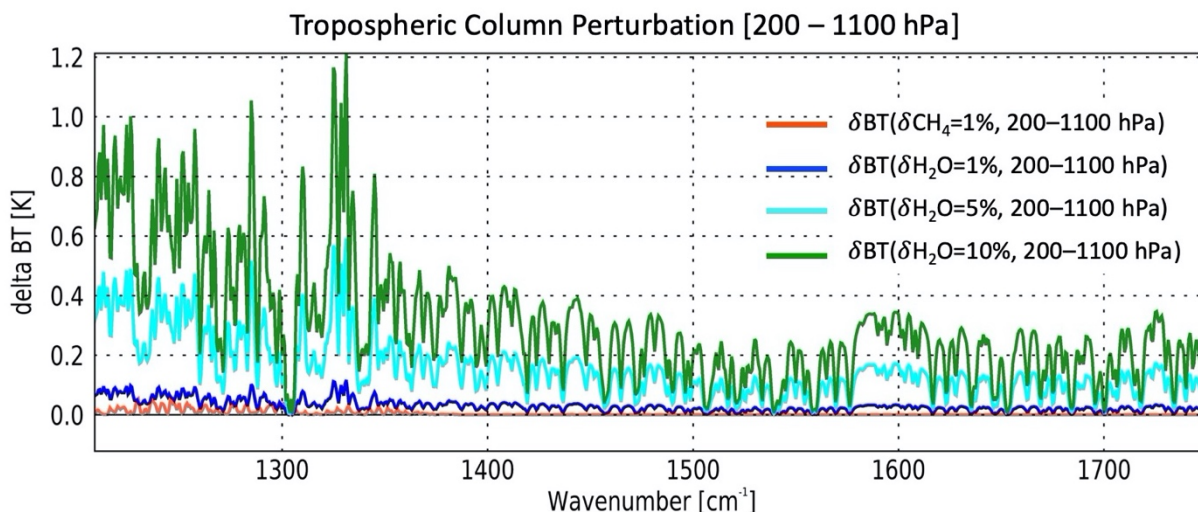


Figure 38: Same as Figure 37 but for absolute values of Brightness Temperature (BT) differences of CH₄ and H₂O across a single tropospheric layer (200–1100 hPa), with varying perturbation values for H₂O. Red: 1% CH₄ perturbation. Blue: 1% H₂O perturbation. Cyan: 5% H₂O perturbation. Green: 10% H₂O perturbation.

Even without such obvious spectral overlap as seen for CH₄ and H₂O vapor (Figures 37 and 38), it is difficult to decompose the highly correlated spectral measurement into distinct atmospheric variables. Space-based observation is especially challenging when you design a system like CLIMCAPS to retrieve a set of 9 distinct profile variables for all weather and climate conditions globally. In CLIMCAPS we adopted a sequential retrieval approach as well as uncertainty quantification and propagation to enable robust retrievals of multiple atmospheric variables across a wide range of conditions (Smith and Barnett, 2019) but, even then, CH₄ retrieval remains challenging. The use of MERRA-2 as a-priori for temperature and H₂O helps CLIMCAPS mitigate the cross-correlation between CH₄ and H₂O vapor in the mid-wave IR band. Other sequential optimal estimation (OE; Rodgers, 2000) systems, such as AIRS V6 or NUCAPS, use statistical regressions implemented either as neural networks or linear retrievals that use all IR and microwave spectral channels. When a system uses spectral channels twice in a retrieval (first to calculate its a-priori, and then to calculate its Bayesian solution), the strong spectral correlations are amplified and propagated. This is one of the reasons why you see large spatial variation in NUCAPS (or AIRS v.7) CH₄ retrievals. Another reason is that both AIRS v.7 and NUCAPS iterate cloud clearing and thus propagate errors due to cloud estimation and removal through the regression retrieval into the final solution. We discuss the benefits of using MERRA 2 as CLIMCAPS a-priori for temperature and H₂O in Table 2 of Smith and Barnett (2019).

Efforts to globally monitor CH₄ sources, sinks and trends are stifled by a lack of measurements in general and where measurements do exist large inconsistencies prevail (Duren et al., 2019). Some space-borne instruments do show potential for monitoring CH₄ globally with reports on CH₄ source detection from TROPOMI (Hu et al., 2018; Pandey et al., 2019) and Hyperion on EO-1 (Thompson et al., 2016), both instruments with high spectral resolution short-wave infrared capability (~2.3 μm) and strong sensitivity to CH₄ molecules in the boundary layer.

We anticipate that CLIMCAPS CH₄ products may add value when used in conjunction with other data sources, such as those from TROPOMI, to add synoptic-scale information about mid- to upper-tropospheric CH₄ patterns under certain conditions. For example, CLIMCAPS CH₄ retrievals could compliment the monitoring of Amazonian CH₄ (Bloom et al., 2016) or CH₄ changes in the Arctic (Shakhova et al., 2010; Thornton et al., 2016).

Figure 39 depicts full-spectral resolution CLIMCAPS-SNPP CH₄ retrievals as tropospheric column densities [molec/cm²] integrated over pressure layers between 200 and 700 hPa for all ascending orbits (13h30 local overpass time) on 1 April 2016. We binned the CLIMCAPS CH₄ column densities to an equal-angle 1.5° grid and averaged all retrievals that passed quality control within each grid cell. From the netCDF Level 2 product file, we access the CLIMCAPS profile retrievals of CH₄ in the **mol_lay/ch4_mol_lay** field, and quality control metrics in **mol_lay/ch4_mol_lay_qc**. We integrated and averaged all values where **mol_lay/ch4_mol_lay_qc** = 0 (i.e., best quality). Figure 39 closely resembles the spatial variability of Level 2 CH₄ retrievals since CLIMCAPS footprints at edge of scan in the Tropics is ~1.5° and at nadir ~0.5°.

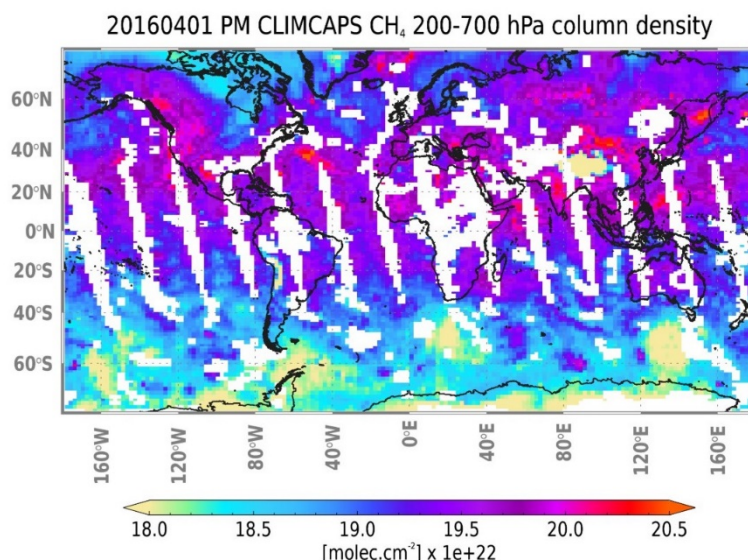


Figure 39: CLIMCAPS-SNPP CH₄ column density retrievals [molec/cm²] from full-spectral resolution CrIS integrated over mid- to upper tropospheric pressure layers between 200 and 700 hPa. The CLIMCAPS CH₄ column densities were gridded and averaged on a 1.5-degree global grid to closely resemble variation at native product resolution in the Level 2 files. Note that the color scale is amplified to highlight spatial patterns **mol_lay/ch4_mol_lay** field, and quality control metrics in **mol_lay/ch4_mol_lay_qc**. We integrated and averaged all values where **mol_lay/ch4_mol_lay_qc** = 0 (i.e., best quality). Much of the variability in CH₄ visible here can be attributed to variability in H₂O vapor, another gas species radiatively active in the 1300cm⁻¹ spectral range from which CLIMCAPS select its CH₄ channels for retrieval. Compared to CH₄, H₂O has a stronger spectral signal which complicates the separation of these two gases during retrieval.

Overall, CLIMCAPS has lower values for CH₄ in the Southern Hemisphere (Figure 39), which can partly be attributed to the a-priori that has a distinct North-South gradient. Much of the daily tropospheric variability in CH₄ visible in Figure 39 can, however, be attributed to variability in H₂O vapor, which CLIMCAPS treats as an interfering gas and major source of uncertainty in its CH₄ retrievals. CLIMCAPS retrieves CH₄ only after it retrieved H₂O vapor at the target scene to minimize this interference.

Figure 40 depicts CLIMCAPS-SNPP information content for CH₄ using degrees of freedom (DOF) as metric. Here we do not see a North-South gradient, but instead a similar pattern of observability in the Northern and Southern Hemispheres. This is partly due to the fact that, unlike CH₄ column densities (Figure 39), the DOF (Figure 40) contains no explicit knowledge of the CLIMCAPS CH₄ a-priori.

Overall, full spectral resolution CLIMCAPS-SNPP has a DOF of ~0.7 for CH₄ with values rarely exceeding 1.0 and sometimes falling below 0.4. This means that CLIMCAPS-SNPP generally has a weak ability to observe a tropospheric layer of CH₄ (i.e., when DOF = 1.0) and most of the time contribute some information only to a partial column. CLIMCAPS CH₄ information content is similar for all systems; CLIMCAPS-SNPP with normal spectral resolution CrIS, CLIMCAPS-NOAA20 and CLIMCAPS-Aqua.

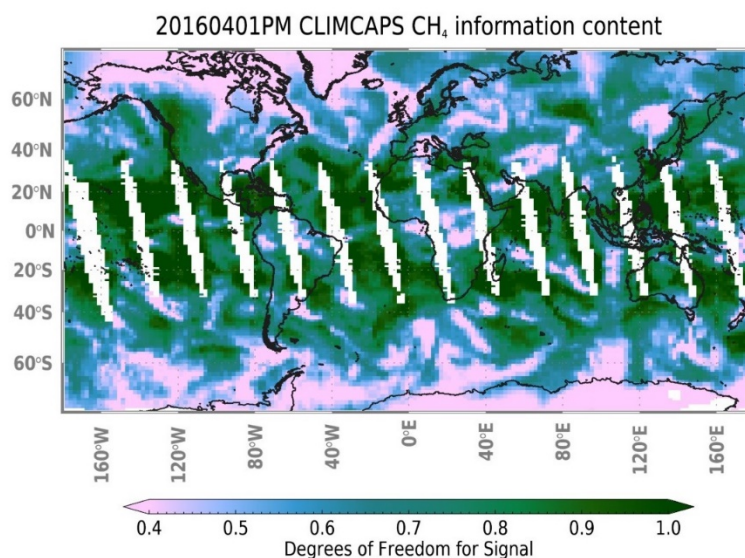


Figure 40: CLIMCAPS-SNPP degrees of freedom (DOF) for CH₄ from full-spectral resolution CrIS at every retrieval scene from ascending orbits (13h30 local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about CO at every scene. For most of the globe, CLIMCAPS has CH₄ DOF < 1. We used the netCDF field **ch4_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome.

We can diagnose CLIMCAPS along the vertical atmospheric pressure gradient. Figure 41 depicts mean profiles (with standard deviation error bars) of (i) sensitivity to the true state given by the averaging kernel matrix diagonal, (ii) layer density retrievals and (iii) a-posteriori error for the North Polar region (> 60°North). Figure 42 is similar to Figure 41 but for the Tropics (30°South

to 30°North). Note how there are fewer vertical error bars on the averaging kernel profile (blue line) compared to the retrieval (orange line) and error (yellow line) profiles. This is because CLIMCAPS averaging kernels are calculated on a reduced set of pressure layers, defined by a series of overlapping trapezoids.

Note how the averaging kernels (blue lines) show that CLIMCAPS has maximum sensitivity to CH₄ broadly around 400 hPa at high latitudes (Figure 41) and ~300 hPa at low latitudes (Figure 42).

The retrieval error profiles (yellow line) in Figures 41 and 42 represent the diagonal vector of the a-posteriori error covariance matrix that CLIMCAPS generates for each retrieval variable at each field of regard. This error profile in the netCDF file (**mol_lay/co_mol_lay_err**) has the same units [molec/cm²] as the retrieval profile (**mol_lay/co_mol_lay**) so we could easily calculate the error as a percentage by dividing the error by the retrieval, multiplied by 100. This error does *not* represent the accuracy, bias, or error with respect to the true state, but instead is a representation of how much CLIMCAPS improved upon the a-priori error estimate, given system uncertainty.

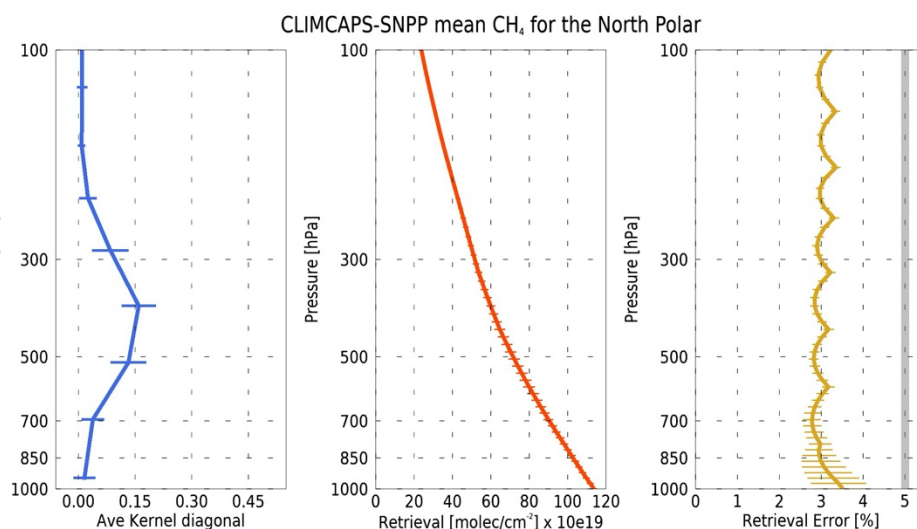


Figure 41: A diagnosis of CLIMCAPS-SNPP CH₄ retrievals for the North Polar latitudinal zone [>60°N] on 1 April 2016. Each solid line represents the mean zonal profile and error bars the standard deviation at each pressure layer. [left] CLIMCAPS CH₄ averaging kernel matrix diagonal vector from netCDF field **ave_kern/ch4_ave_kern** that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of CH₄ in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field **mol_lay/ch4_mol_lay** [molec/cm²]. [right] CLIMCAPS retrieval error from netCDF field **mol_lay/ch4_mol_lay_err** [molec/cm²] represented here as percentage [**mol_lay/ch4_mol_lay_err**]/[**mol_lay/ch4_mol_lay**]*100. CLIMCAPS uses a CH₄ a-priori error of 5% as represented by the thick grey line in panel on the right. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals. In calculating these mean profiles, we filtered out all retrievals where **mol_lay/ch4_mol_lay_qc**(*,*i,j*) ≥ 1. We plot these profiles using the pressure layer array from **air_pres_lay***100 in hPa units.

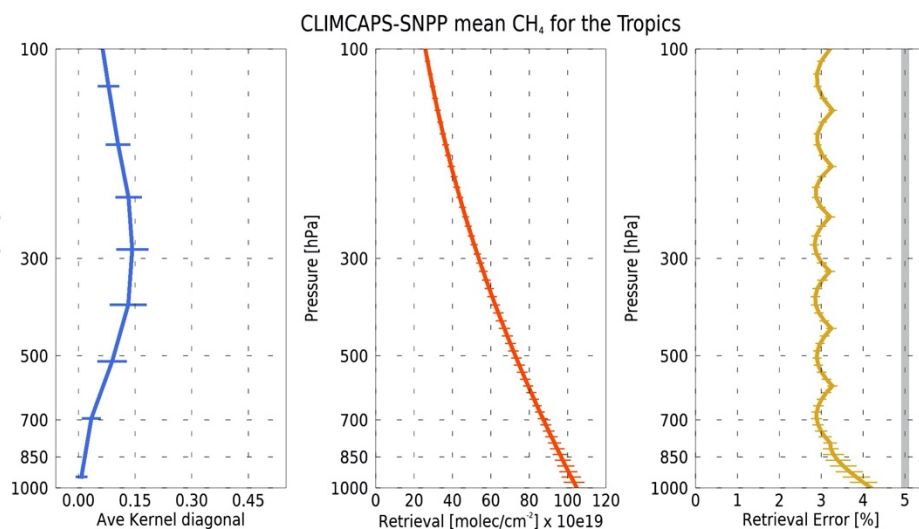


Figure 42: Same as Figure 41 but for the Tropical zone [30°S to 30°N].

CLIMCAPS defines a 5% error for its CH₄ a-priori, which we indicate with the thick grey line in the righthand panels of Figures 41 and 42. Here we see that CLIMCAPS reduces the a-priori error to ~3% within the pressure layers defined by the averaging kernel. The retrieval error is a typical by-product of all systems using the Rodgers (2000) OE retrieval method and should be used as a metric for characterizing the observing system, not as a metric for defining the accuracy or bias of the retrieval itself (Smith and Barnett, 2019).

5. Preparing CH₄ retrievals for applications

– Boundary layer adjustment

CLIMCAPS uses a standard 100-layer pressure grid to retrieve atmospheric variables from Earth surface (1100 hPa) to top of atmosphere (0.005 hPa). This pressure grid is required by radiative transfer models (SARTA for CLIMCAPS) to accurately calculate top of atmosphere hyperspectral IR radiances. CLIMCAPS uses the exact same pressure grid at every scene on Earth and accounts for surface pressure as a separate variable during radiative transfer calculations. The retrieved profiles are, however, reported on the 100-layer grid as a means to standardize the output. It is important that you adjust the bottom layer, i.e. that pressure layer intersecting the Earth surface as identified by `air_pres_lay_nsurf` in the CLIMCAPS netCDF file, to accurately reflect the total number of CH₄ molecules in the boundary layer.

This boundary layer adjustment is important if you calculate total column CH₄ densities and we describe the method in Chapter 3, Section 4.

This boundary layer adjustment is irrelevant if you work with CH₄ retrievals in the mid-troposphere, above the boundary layer such as depicted in Figure 39.

– *Space-time averaging*

CLIMCAPS CH₄ retrievals must be aggregated in space and time before application to remove random uncertainty, specifically H₂O vapor. Researchers at the NOAA Earth Systems Research Lab (ESRL) performed an in-depth study of the degree to which CH₄ retrieved from space-borne IR systems should be aggregated using scale variance techniques (Frost et al., 2018; McKeen et al., 2016). For retrievals of CH₄ from the NOAA-Unique Combined Atmospheric Processing System (NUCAPS) for CrIS on SNPP and NOAA-20 (a sister system to CLIMCAPS) Frost et al., (2018) and McKeen et al., (2016) determined that CH₄ must be aggregated to spatial scales greater than ~340 km (which can roughly be translated to ~3.4° in the Tropics). A report describing their analysis and results is available upon request (Greg Frost, Gregory.J.Frost@noaa.gov, <https://esrl.noaa.gov/csd/staff/gregory.j.frost/>).

Below we demonstrate the value of space-time aggregation for CLIMCAPS CH₄ retrievals by (i) increasing the grid size to 4.0 degrees lat/lon (Figure 43) and (ii) averaging a month of daily (ascending and descending orbits) CLIMCAPS CH₄ observations (Figure 44).

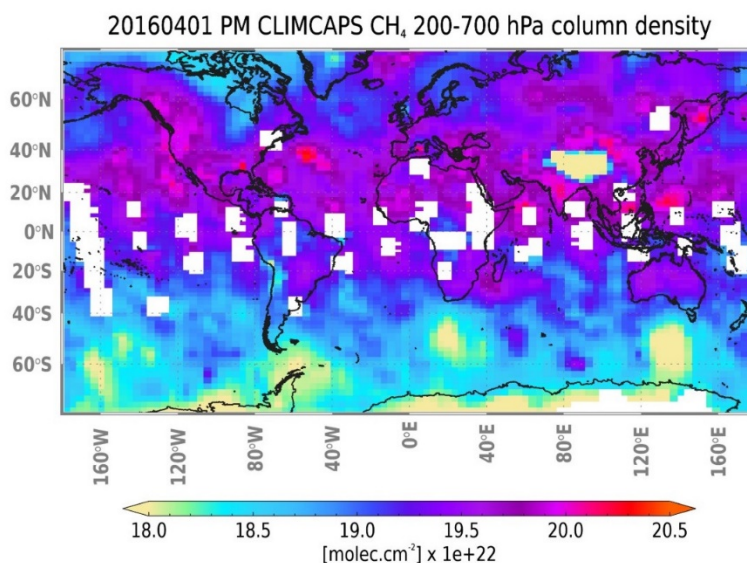


Figure 43: Same as Figure 39 but with values spatially aggregated over a 4° equal-angle grid.

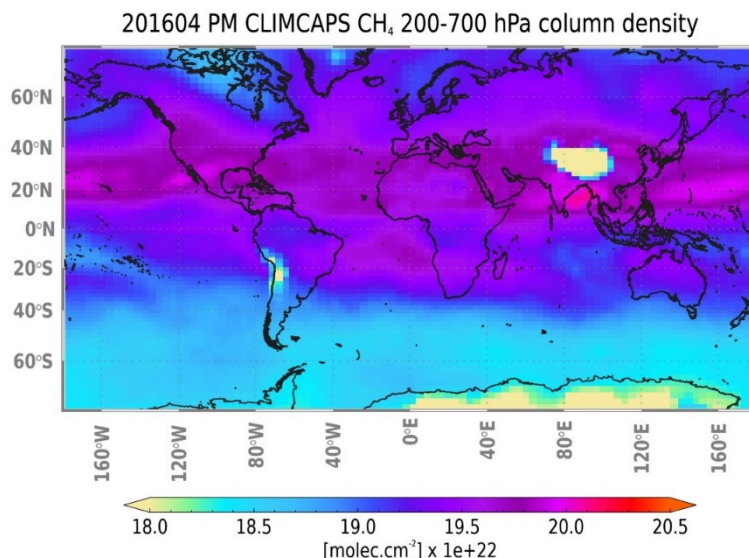


Figure 44: Same as Figure 39 but with values temporally aggregated over a 4° equal-angle grid for a month of retrievals, ascending (13h30 local overpass time) and descending orbits (01h30 local overpass time), 1–30 April 2016.

6. CH₄ a-priori

The CLIMCAP CH₄ a-priori (Figure 45) is derived from performing a non-linear fit based on coefficients trained on in-situ measurements (< 350 hPa) and chemistry model fields (> 350 hPa) from 2015. The coefficients were calculated off-line as described by Xiong et al., (2008, 2013) and the a-priori profiles are calculated for the retrieval scene's latitude at run-time. Figure 45 depicts the CLIMCAPS CH₄ a-priori as mixing ratio [ppb] across all latitudes, with higher values in the Northern Latitudes below 300 hPa. The greatest variability in CH₄ occurs in the mid-troposphere with a large drop off above the tropopause. Unlike the CLIMCAPS a-priori's for other retrieval variables, there is no temporal or longitudinal variability in the CH₄ a-priori. therefore, any temporal or longitudinal structure in the CLIMCAPS methane product is derived from the observation itself. CLIMCAPS uses a 5% error for CH₄ a-priori at all retrieval scenes.

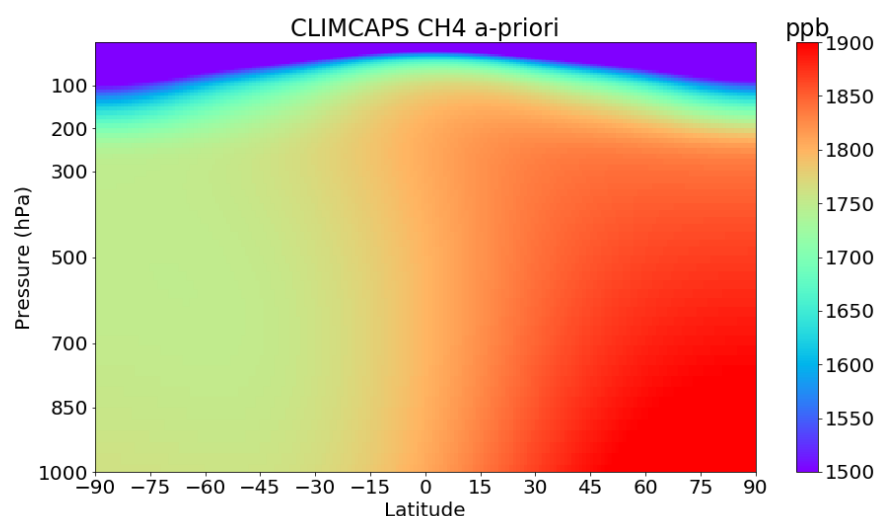


Figure 45: CLIMCAPS CH₄ a-priori mixing ratio [ppb] profiles calculated along each latitude from North to South. The CLIMCAPS CH₄ a-priori is calculated from a set of coefficients and varies with pressure [hPa] and latitude. These coefficients were developed by Xiong et al.(2008, 2013).

Improving the CLIMCAPS a-priori for CH₄ retrieval is an area of ongoing research.

7. CLIMCAPS product field names relevant to CH₄ applications

Within the netCDF files, the following fields are relevant for CH₄ studies. Each CLIMCAPS file contains 45 scanlines along track (atrack) and 30 FOR along each scanline, or across track (xtrack). With CH₄ profiles retrieved at each FOR on 100 pressure layers (air_pres_layer), the arrays have dimensions [atrack, xtrack, air_pres_layer].

– Retrieved variables

- **mol_layer/ch4_mol_layer**: This is the integrated column amount of CH₄ from the top of the atmosphere (TOA = 0.005 hPa) to the surface. Note that the retrieval is typically only sensitive around 500 hPa, so this should be combined with the averaging kernel (ch4_ave_kern) to understand how this field differs from the first guess.
- **mol_layer/ch4_mol_layer_err co_mol_layer** error estimate.

– Quality metrics

- **mol_layer/ch4_mol_layer_qc**: profile quality control metrics ranging from 0 = good, 1 = suspect, 2 = bad.
- **ave_kern/ch4_ave_kern**: CH₄ averaging kernel.
- **ch4_dof**: The trace of the averaging kernel matrix as a measure of the number of pieces of information about the methane profile provided by the physical retrieval step. Degrees of freedom indicate the number of distinct vertical levels that the algorithm has sensitivity. For CH₄, this is typically below 1.0.

References

- AAAS. Stratospheric polar vortex influences winter cold, researchers say. Retrieved April 7, 2020, from http://www.eurekalert.org/pub_releases/2001-12/uoi-a-spv120301.php, 2001.
- Berndt, E. and Folmer, M.: Utility of CrIS/ATMS profiles to diagnose extratropical transition, *Results in Physics*, 8, 184–185, doi:10.1016/j.rinp.2017.12.006, 2018.
- Bloom, A. A., Lauvaux, T., Worden, J., Yadav, V., Duren, R., Sander, S. P. and Schimel, D. S.: What are the greenhouse gas observing system requirements for reducing fundamental biogeochemical process uncertainty? Amazon wetland CH₄ emissions as a case study, *Atmospheric Chemistry and Physics*, 16(23), 15199–15218, doi:10.5194/acp-16-15199-2016, 2016.
- Cooper, O. R., Parrish, D. D., Ziemke, J., Balashov, N. V., Cupeiro, M., Galbally, I. E., Gilge, S., Horowitz, L., Jensen, N. R., Lamarque, J.-F., Naik, V., Oltmans, S. J., Schwab, J., Shindell, D. T., Thompson, A. M., Thouret, V., Wang, Y. and Zbinden, R. M.: Global distribution and trends of tropospheric ozone: An observation-based review, *Elementa: Science of the Anthropocene*, 2, 000029, doi:10.12952/journal.elementa.000029, 2014.
- Dessler, A. E. and Wong, S.: Estimates of the Water Vapor Climate Feedback during El Niño–Southern Oscillation, *Journal of Climate*, 22(23), 6404–6412, doi:10.1175/2009JCLI3052.1, 2009.
- Dreessen, J.: A Sea Level Stratospheric Ozone Intrusion Event Induced within a Thunderstorm Gust Front, *Bulletin of the American Meteorological Society*, 100(7), 1259–1275, doi:10.1175/BAMS-D-18-0113.1, 2019.
- Drummond, J. R. and Mand, G. S.: The Measurements of Pollution in the Troposphere (MOPITT) Instrument: Overall Performance and Calibration Requirements, *Journal of Atmospheric and Oceanic Technology*, 13, 314–320, [https://doi.org/10.1175/1520-0426\(1996\)013<0314:TMOPIT>2.0.CO;2](https://doi.org/10.1175/1520-0426(1996)013<0314:TMOPIT>2.0.CO;2), 1996.
- Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O. and Miller, C. E.: California's methane super-emitters, *Nature*, 575(7781), 180–184, doi:10.1038/s41586-019-1720-3, 2019.
- Frost, G. J., McKeen, S. A., Smith, N., Barnet, C. D., Ahmadov, R., Pierce, R. B., Ryerson, T. B., Holloway, J. S., Peischl, J., Pollack, I. B., McKain, K., Sweeney, C., Daube, B. C., Commane, R., Wofsy, S. C., Thompson, C. R. and Bourgeois, I.: Characterizing Tropospheric Trace Gas Retrievals from the Cross-track Infrared Sounder, in *AGU Fall Meeting Abstracts*, vol. 2018, pp. IN44B-06., 2018.
- Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados-Muñoz, M. J., Hannigan, J. W., Hase, F., Hassler, B., Huang, G., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S., Latter, B., Leblanc, T., Le Flochmoën, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M., Trickl, T., Weatherhead, E., Wespes, C., Worden, H. M., Vigouroux, C., Xu, X., Zeng, G., Ziemke, J., Helmig, D. and Lewis, A.: Tropospheric

- Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, *Elem Sci Anth*, 6, doi:10.1525/elementa.291, 2018.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *Journal of Climate*, 30(14), 5419–5454, doi:10.1175/JCLI-D-16-0758.1, 2017.
- GMAO: MERRA-2 inst3_3d_asm_Nv: 3d,3-Hourly, Instantaneous,Model-Level,Assimilation,Assimilated Meteorological Fields V5.12.4, , doi:10.5067/WWQSQ8IVFW8, 2015.
- Gray, E.: 2019 Ozone Hole Is the Smallest on Record Since Its Discovery, *The Earth Observer*, 31(6), 34–35, 2019.
- Hu, H., Landgraf, J., Detmers, R., Borsdorff, T., Aan de Brugh, J., Aben, I., Butz, A. and Hasekamp, O.: Toward Global Mapping of Methane With TROPOMI: First Results and Intersatellite Comparison to GOSAT, *Geophysical Research Letters*, 45(8), 3682–3689, doi:10.1002/2018GL077259, 2018.
- Iturbide-Sanchez, F., da Silva, S. R. S., Liu, Q., Pryor, K. L., Pettey, M. E. and Nalli, N. R.: Toward the Operational Weather Forecasting Application of Atmospheric Stability Products Derived From NUCAPS CrIS/ATMS Soundings, *IEEE Transactions on Geoscience and Remote Sensing*, 56(8), 4522–4545, doi:10.1109/TGRS.2018.2824829, 2018.
- Langford, A. O., Alvarez, R. J., Brioude, J., Evan, S., Iraci, L. T., Kirgis, G., Kuang, S., Leblanc, T., Newchurch, M. J., Pierce, R. B., Senff, C. J. and Yates, E. L.: Coordinated profiling of stratospheric intrusions and transported pollution by the Tropospheric Ozone Lidar Network (TOLNet) and NASA Alpha Jet experiment (AJAX): Observations and comparison to HYSPLIT, RAQMS, and FLEXPART, *Atmospheric Environment*, 174, 1–14, doi:10.1016/j.atmosenv.2017.11.031, 2018.
- Maddy, E. S. and Barnet, C. D.: Vertical Resolution Estimates in Version 5 of AIRS Operational Retrievals, *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2375–2384, doi:10.1109/TGRS.2008.917498, 2008.
- Maddy, E. S. and Barnet, C. D.: Vertical resolution estimates in Version 5 of AIRS operational retrievals, 46, 2375–2384, <https://doi.org/10.1109/TGRS.2008.917498>, 2008.
- Maddy, E. S., Barnet, C. D., and Gambacorta, A.: A computationally efficient retrieval algorithm for hyperspectral sounders incorporating a-priori information, 6, 802–806, <https://doi.org/10.1109/LGRS.2009.2025780>, 2009.
- McKeen, S. A., Angevine, W. M., Ahmadov, R., Frost, G. J., Kim, S. W., Cui, Y., McDonald, B., Trainer, M., Holloway, J. S., Ryerson, T. B., Peischl, J., Gambacorta, A., Barnet, C. D., Smith, N. and Pierce, R. B.: Mid-tropospheric Spectral Length-scale Analysis of Many Constituents from Aircraft, Satellite and Model Results During the 2013 SENEX Field Study., in *AGU Fall Meeting Abstracts*, vol. 2016, pp. IN31A-1746., 2016.

- Nalli, N. R., Gambacorta, A., Liu, Q., Barnet, C. D., Tan, C., Iturbide-Sanchez, F., Reale, T., Sun, B., Wilson, M., Borg, L. and Morris, V. R.: Validation of atmospheric profile retrievals from the SNPP NOAA-Unique Combined Atmospheric Processing System. Part 1: Temperature and Moisture, *IEEE Transactions on Geoscience and Remote Sensing*, 56(1), 180–190, doi:10.1109/TGRS.2017.2744558, 2018a.
- Nalli, N. R., Gambacorta, A., Liu, Q., Tan, C., Iturbide-Sanchez, F., Barnet, C. D., Joseph, E., Morris, V. R., Oyola, M. and Smith, J. W.: Validation of atmospheric profile retrievals from the SNPP NOAA-Unique Combined Atmospheric Processing System. Part 2: Ozone, *IEEE Transactions on Geoscience and Remote Sensing*, 56(1), 598–607, doi:10.1109/TGRS.2017.2762600, 2018b.
- Pandey, S., Gautam, R., Houweling, S., van der Gon, H. D., Sadavarte, P., Borsdorff, T., Hasekamp, O., Landgraf, J., Tol, P., van Kempen, T., Hoogeveen, R., van Hees, R., Hamburg, S. P., Maasackers, J. D. and Aben, I.: Satellite observations reveal extreme methane leakage from a natural gas well blowout, *Proceedings of the National Academy of Sciences*, 116(52), 26376–26381, doi:10.1073/pnas.1908712116, 2019.
- Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore; Hackensack, N.J., 2000.
- Rosenkranz, P. W. and Barnet, C. D.: Microwave radiative transfer model validation, *Journal of Geophysical Research*, 111(D9), doi:10.1029/2005JD006008, 2006.
- Rosenkranz, P. W.: Cloud liquid-water profile retrieval algorithm and validation, *Journal of Geophysical Research*, 111(D9), doi:10.1029/2005JD005832, 2006.
- Rosenkranz, P. W.: Rapid radiative transfer model for AMSU/HSB channels, *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 362–368, doi:10.1109/TGRS.2002.808323, 2003.
- Rosenkranz, P. W.: Retrieval of temperature and moisture profiles from AMSU-A and AMSU-B measurements, *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), 2429–2435, doi:10.1109/36.964979, 2001.
- Shakhova, N., Semiletov, I., Salyuk, A., Yusupov, V., Kosmach, D. and Gustafsson, O.: Extensive Methane Venting to the Atmosphere from Sediments of the East Siberian Arctic Shelf, *Science*, 327(5970), 1246–1250, doi:10.1126/science.1182221, 2010.
- Smith, N. and Barnet, C. D.: CLIMCAPS Observing Capability for Temperature, Moisture and Trace Gases from AIRS/AMSU and CrIS/ATMS, preprint, *Gases/Remote Sensing/Data Processing and Information Retrieval*, 2020.
- Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), *Remote Sensing*, 11(10), 1227, doi:10.3390/rs11101227, 2019.
- Strow, L. L., Hannon, S. E., De Souza-Machado, S., Motteler, H. E. and Tobin, D.: An overview of the AIRS radiative transfer model, *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 303–313, doi:10.1109/TGRS.2002.808244, 2003.
- Té, Y., Jeseck, P., Franco, B., Mahieu, E., Jones, N., Paton-Walsh, C., Griffith, D. W. T., Buchholz, R. R., Hadji-Lazaro, J., Hurtmans, D., and Janssen, C.: Seasonal variability of surface and column carbon monoxide over the megacity Paris, high-altitude Jungfraujoch and Southern Hemispheric Wollongong stations, 16, 10911–10925, <https://doi.org/10.5194/acp-16-10911-2016>, 2016.

- Thompson, D. R., Thorpe, A. K., Frankenberg, C., Green, R. O., Duren, R., Guanter, L., Hollstein, A., Middleton, E., Ong, L. and Ungar, S.: Space-based remote imaging spectroscopy of the Aliso Canyon CH₄ superemitter: ORBITAL SPECTROSCOPY OF THE ALISO EVENT, *Geophysical Research Letters*, 43(12), 6571–6578, doi:10.1002/2016GL069079, 2016.
- Thornton, B. F., Geibel, M. C., Crill, P. M., Humborg, C. and Mörtz, C.-M.: Methane fluxes from the sea to the atmosphere across the Siberian shelf seas, *Geophysical Research Letters*, 43(11), 5869–5877, doi:10.1002/2016GL068977, 2016.
- Wargan, K., Kramarova, N., Weir, B., Pawson, S. and Davis, S. M.: Toward a Reanalysis of Stratospheric Ozone for Trend Studies: Assimilation of the Aura Microwave Limb Sounder and Ozone Mapping and Profiler Suite Limb Profiler Data, *Journal of Geophysical Research: Atmospheres*, 125(4), doi:10.1029/2019JD031892, 2020.
- Wargan, K., Labow, G., Frith, S., Pawson, S., Livesey, N. and Partyka, G.: Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis, *Journal of Climate*, 30(8), 2961–2988, doi:10.1175/JCLI-D-16-0699.1, 2017.
- Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L. and Emma Knowland, K.: Recent Decline in Extratropical Lower Stratospheric Ozone Attributed to Circulation Changes, *Geophysical Research Letters*, 45(10), 5166–5176, doi:10.1029/2018GL077406, 2018.
- Witze, A.: Rare ozone hole opens over Arctic – and it's big, *Nature*, 580(7801), 18–19, doi:10.1038/d41586-020-00904-w, 2020.
- Xiong, X., Barnett, C., Maddy, E. S., Gambacorta, A., King, T. S. and Wofsy, S. C.: Mid-upper tropospheric methane retrieval from IASI and its validation, *Atmospheric Measurement Techniques*, 6(9), 2255–2265, doi:10.5194/amt-6-2255-2013, 2013.
- Xiong, X., Barnett, C., Maddy, E., Sweeney, C., Liu, X., Zhou, L. and Goldberg, M.: Characterization and validation of methane products from the Atmospheric Infrared Sounder (AIRS), *Journal of Geophysical Research*, 113(null), doi:10.1029/2007JG000500, 2008.

Chapter 3: CLIMCAPS Levels and Layers

Satellite soundings, unlike satellite imagery, provide estimates of temperature water vapor (H_2O) and trace gases as vertical profiles along a pressure grid [hPa], from the Earth surface to top of atmosphere. CLIMCAPS reports its retrieved soundings on the standard 100 level pressure grid also used in SARTA (Strow et al., 2003), which spans 1100 hPa to 0.016 hPa. CLIMCAPS retrieves $T(p)$ as the temperature at each pressure *level*, while it retrieves H_2O , O_3 , CO , CH_4 , HNO_3 , N_2O and SO_2 (i.e., all gases except CO_2) as column densities [molec/cm^2] within pressure *layers*. A pressure layer is the vertical area between two pressure *levels*. CLIMCAPS retrieves CO_2 as pressure layer mixing ratio [ppb], although CO_2 is reported in the product files as volume mixing ratio on pressure levels, which is the mole of CO_2 per mole of dry air [mol/mol].

In this document we discuss issues that can significantly affect data application and understanding as they relate to CLIMCAPS retrieval levels and layers. These are:

1. Levels versus Layers, and converting from one to the other.
2. Adjusting retrieved profiles to scene surface pressure

Section 1: Levels versus Layers

CLIMCAPS requires all state variables to be on the standard 100 pressure levels used in SARTA (Strow et al., 2003) for accurate calculations of top of atmosphere infrared radiance in narrow spectral intervals. This is a fixed 100-level pressure grid (P_{lev}) and available in the CLIMCAPS product file as **air_pres**, which is the same at all scenes across the globe irrespective of variations in Earth surface pressure. $P_{\text{lev}}(1) = 0.016$ hPa, and $P_{\text{lev}}(100) = 1100.0$ hPa.

CLIMCAPS does not retrieve surface pressure but instead uses MERRA2 surface pressure as state variable and write it to the product files as: **aux/prior_surf_pres** [hPa]. The MERRA2 surface pressure values written to **aux/prior_surf_pres** [hPa] have been spatially and temporally interpolated to the CLIMCAPS footprint.

Over the ocean, surface pressure is typically 1013.25 hPa. High terrain surface pressure is often less than 800 hPa, such as over the Tibetan Plateau in Asia, the Rocky Mountains in North America, and the Andes Mountains in South America. Most of the time $P_{\text{lev}}(100)$ is well below Earth surface pressure, which means that one needs to truncate the retrieved profiles according to the true Earth surface pressure at the target scene. The CLIMCAPS product field, **air_pres_nsurf**, contains the index value where the standard **air_pres** and **air_pres_lay** arrays intersects with surface pressure (**aux/prior_surf_pres**) at a target retrieval scene. In each CLIMCAPS product file there are **air_pres** and **air_pres_lay** with dimension [*air_pres*] where $\text{air_pres} = 100$, and **air_pres_nsurf** with dimension [*atrack* x *xtrack*], where *atrack* = 30 to indicate the number of retrieval footprints along each scanline and *xtrack* = 45 to indicate the number of scanlines per file. One should truncate temperature retrieval at position *atrack* = *i* and *xtrack* = *j* with $P_{\text{bot}} = \text{air_pres_nsurf}(i, j)$ such that $T(p) = \text{air_temp}(i, j, 1:P_{\text{bot}})$. Stated differently, CLIMCAPS profile retrievals are valid from the top of atmosphere, **air_pres**(1), to

the Earth surface as defined by MERRA2 surface pressure (**aux/prior_surf_pres**) and indicated by **air_pres_nsurf** for each retrieval scene

CLIMCAPS retrieves trace gas quantities on a standard pressure layer grid (**air_pres_lay**), which is a 100 layer grid derived from **air_pres**. Figure 46 below gives a graphic representation of *levels* versus *layers*.

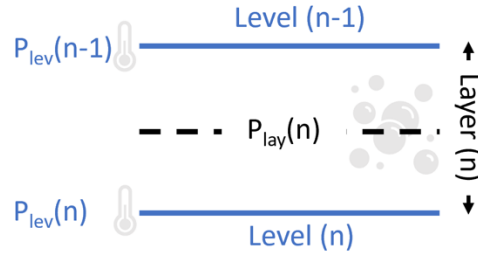


Figure 46: An illustration showing the differences between CLIMCAPS pressure levels (P_{lev}) [**air_pres**] and pressure layers (P_{lay}) [**air_pres_nsurf**] for an index n , where $n \in \{1 \dots 100\}$. Temperature is retrieved on levels while trace gasses are retrieved on pressure layers. The mathematical conversion from levels to layers is shown in equation 6.

We derive the pressure layer (P_{lay}) grid from pressure levels (P_{lev}) as follows:

$$P_{lay}(n) = [P_{lev}(n) - P_{lev}(n-1)] / [\log_{10} \left(\frac{P_{lev}(n)}{P_{lev}(n-1)} \right)] \quad 6$$

where $n \in \{2 \dots 100\}$, $P_{lay}(1) = 0.00284$ hPa and $P_{lay}(100) = 1085.36$ hPa. All trace gas quantities in the **mol_lay** group, as well as **aux/co2_vmr** are retrieved and reported on this pressure *layer* grid. It is thus important that you associate the CLIMCAPS trace gas retrieval profiles with pressure layers (**air_pres_lay**) not levels (**air_pres**).

Section 2: Converting Levels to Layers

Some applications may require you to evaluate CLIMCAPS temperature retrieval levels alongside H₂O density layers. In such cases, convert **air_temp** from its native definition on **air_pres** *levels* to **air_pres_lay** *layers* as follows:

Assume that the top of atmosphere temperature is isothermal (Eq. 7), while all other layers are calculated as two-point running means (Eq. 8).

$$T_{lay}(n) = T_{lev}(n), n = 1 \quad 7$$

$$T_{lay}(n) = 0.5(T_{lev}(n) + T_{lev}(n-1)), n \in \{2 \dots 100\} \quad 8$$

Section 3: Converting Layers to Levels

If you wish to represent CLIMCAPS gas quantities on pressure levels, we recommend the procedure below. It is important that you start off of the profiles in column density units

[molecules/cm²] since it ensures that the number of retrieved molecules are preserved. Note that with this conversion from pressure layers to levels we will use water vapor (**h2o_vap**) as example in the pseudocode below.

```
# (1) Extract relevant fields from the netCDF file
Play = file["air_pres_layer"]
Plev = file["air_pres"]
h2olay = file["mol_layer/h2o_vap_mol_layer"]

# (2) Calculate mid-point between two pressure layers
h2omid(1) = h2olay(1)
h2omid(n) = 0.5*[h2olay(n-1) + h2olay(n)], where n=2,100

# (3) Interpolate to pressure level grid using the natural logarithmic form of the
profiles to linearize the calculation
log_h2olev = INTERPOLATE(log10(h2omid), log10(Play), log10(Plev))
h2olev = 10^(log_h2olev)
```

We strongly discourage the interpolation of mixing ratio or relative humidity from layers to levels as this may introduce an off-set or change the retrieved profile in a way that cannot be quantified.

Section 4: Boundary Layer Adjustment

CLIMCAPS retrieves atmospheric variables along a standard pressure grid that does not account for variation in surface pressure, P_s . In the CLIMCAPS product file we record the index along the standard pressure level (**air_pres**) and pressure layer (**air_pres_layer**) arrays as **air_pres_nsurf** to indicate the standard pressure value closest to P_s . When we consider $T = \text{air_temp}(i, j, 1:n)$, and $P_{\text{bot}} = \text{air_pres_nsurf}(i, j)$. In most cases, we will see **air_pres**(P_{bot}) either larger or smaller than P_s and rarely, if ever, exactly the same. This means one of two scenarios can occur as demonstrated in Figure 47 below.

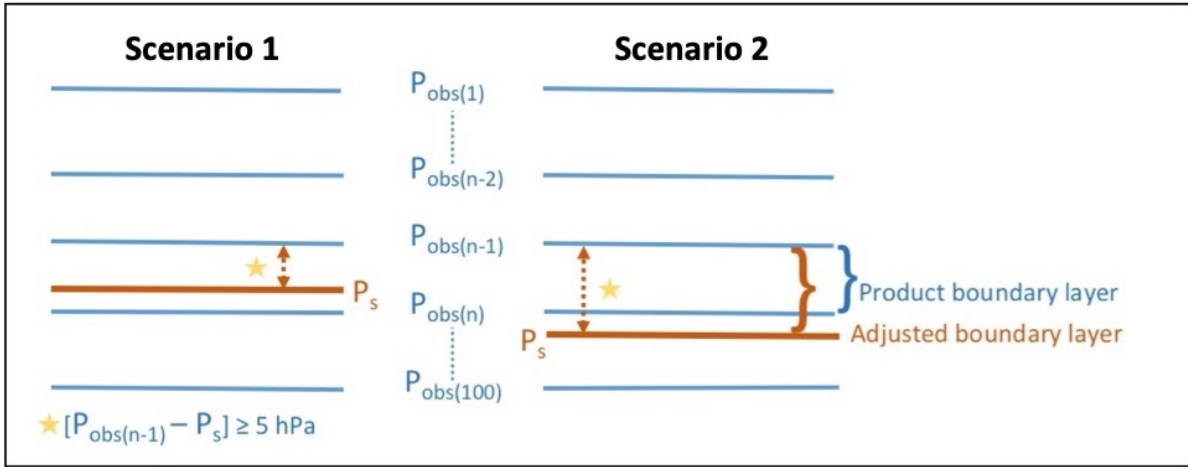


Figure 47: An illustration of the two types of boundary layer conditions that result when retrieving atmospheric variables on a standard pressure grid, P_{obs} , which for CLIMCAPS is the pressure levels profile, **air_pres**. CLIMCAPS uses surface pressure, P_s , from MERRA2. Here the subscript n denotes ranges from 1 to 100, and represents the index at which the CLIMCAPS retrieval is reported at. Unless the retrieval is performed over ocean $P_{obs(n)}$ is always less than $P_{obs(plev)}$ with $n < plev$, $plev = 100$, and $P_{obs(plev)} = 1100$ mb. **Scenario 1** requires a narrowing of the boundary layer with interpolation of the retrieved value from $P_{obs(n-1)}$ to P_s . **Scenario 2** requires a broadening of the boundary layer with extrapolation of the retrieved value from $P_{obs(n-1)}$ to P_s . The brown dotted line indicates the width of the adjusted boundary layer, which always exceeds 5 mb and varies horizontally with surface topography.

We define surface air temperature (T_{surf}) as the atmospheric temperature at surface pressure (P_{surf} , **aux/prior_surf_pres**). We can derive T_{surf} from the retrieved temperature profile (**air_temp**) as follows:

$$T_{surf} = T(P_{bot} - 1) + BL_{mult} \times [T(P_{bot}) - T(P_{bot} - 1)] \quad 9$$

Where BL_{mult} is the boundary layer multiplier that quantifies the degree to which the bottom profile layer needs to be adjusted. P_{bot} (**air_pres_nsurf**) is the index of the pressure level that is closest to P_{surf} while always satisfying the condition:

$$[P_{surf} - P_{lev}(P_{bot})] \geq 5 \text{ hPa} \quad 10$$

This condition ensures that $P_{lev}(P_{bot})$ is always less than P_{surf} by at least 5 hPa. BL_{mult} is given by:

$$BL_{mult} = \frac{P_{surf} - P_{lev}(P_{bot} - 1)}{P_{lev}(P_{bot}) - P_{lev}(P_{bot} - 1)} \quad 11$$

Figure 47 Scenario 1 demonstrates a case where $BL_{mult} < 1.0$, and Scenario 2 a case where $BL_{mult} \geq 1.0$.

Equation [8] convert temperature from level to a layer for any layer in the temperature profile. In context of surface layer, equation [12] can be used to calculate the bottom layer with:

$$T_{lay}(P_{bot}) = \frac{T(P_{bot}-1) + T_{surf}}{2} \quad 12$$

For trace gases, the values are retrieved as *layer column density*, so the boundary layer column density should be adjusted for surface pressure P_{surf} differently than for **air_temp** on pressure levels. The boundary layer for trace gases is indicated by the P_{bot} (**air_pres_nsurf**) of **air_pres_lay**. This bottom layer needs to be either narrowed (Figure 47, Scenario 1) or broadened (Figure 47, Scenario 2) with B_{mult} as follows:

$$CD_{surf} = BL_{mult} \times T_{gas}(P_{bot}) \quad 13$$

Where T_{gas} depicts any of the CLIMCAPS trace gas retrieval species, including CO_2 .

1. Pseudocode for adjusting boundary layer values

Below, we show some example pseudocode to calculate BL_{mult} from equation [11]. Three variables are required from the CLIMCAPS netCDF file: air pressure (**air_pres**), L_{bot} (**air_pres_nsurf**), and P_{surf} (**aux/prior_surf_pres**). P_{surf} is a two-dimensional variable (y (**atrack**), x (**xtrack**)). We simplify our discussion by only calculating BL_{mult} and T_{surf} at the first footprint (**atrack** = **xtrack** = 0).

```
# Extract relevant fields from the netCDF file
pres = file["air_pres"]
Pbot = file["air_pres_nsurf"]
psurf = file["aux/prior_surf_pres"]

# Psurf for the first footprint
psurf0 = psurf[0, 0]
if pres[Pbot] < psurf then Pbot = Pbot - 1

# Calculate the boundary layer multiplier
numerator = psurf0 - pres[Pbot - 1]
denominator = pres[Pbot] - pres[Pbot - 1]
BLmult = numerator / denominator
```

Below we show an example pseudocode to calculate T_{surf} , for a single footprint. We first import the temperature variable (**air_temp**) from the netCDF file. $T(p)$ is a three-dimensional variable (y (**atrack**), x (**xtrack**), and z (**air_pres**)), so like $psurf$ in our BL_{mult} pseudocode above, we keep only the value from the first footprint (**atrack** = **xtrack** = 0). The variable t_{diff} is the difference between the temperature from the level above the bottom level and the temperature from the level at P_{bot} . Surface temperature (t_{surf}) is then calculated on the last line.

```
# Extract relevant fields from the netCDF file
temp[:, :, :] = file["air_temp"]

t_diff = temp[0, 0, Pbot] - temp[0, 0, Pbot - 1])
```

```
t_surf = temp[0, 0, Pbot - 1] + BLmult * t_diff
```

The above code will map a single value to t_{surf} . To calculate T_{surf} over the entire swath, surface temperature can be defined as a two-dimensional vector. Instead of using the first footprint ($\text{atrack} = \text{xtrack} = 0$), the user can write a loop and iterate over atrack and xtrack to calculate the T_{surf} for all footprints in the file(s).

H_2O is often integrated to show column total H_2O . In our example pseudocode below, we import the H_2O variable (**h2o_liq_mol_lay**), and extract the column values for the first footprint ($\text{iobs} = 0$). Then, we compute total H_2O for all layers except the bottom layer, from 1 to $\text{P}_{\text{bot}}-1$. We adjust the bottom layer (P_{bot}) with BL_{mult} and add it to the total, as shown in equation [11]. Like temperature, users can incorporate the procedure from the pseudocode below in a loop to process the whole swath.

```
# Read water vapor column density values from the netCDF file
h2o = file["h2o_liq_mol_lay"]

# Compute total water for all pressure layers up to the layer above the surface layer
total_h2o = sum(h2o[0, 0, i:Pbot-1])

# Then add in the adjusted surface layer quantity
total_h2o = total_h2o + BLmult * h2o [0, 0, Pbot]
```

Chapter 4: Cloud Clearing and Cloud Retrievals

1. Does CLIMCAPS retrieve cloud parameters?

Yes – cloud top pressure (C_{TP}) and cloud fraction (α). CLIMCAPS retrieves cloud fraction for each field-of-view (FOV; instrument footprint) and a two-layer cloud top pressure for each field-of-regard (FOR; CLIMCAPS footprint). There are nine FOVs in each FOR (see Figure 1 below). This means that CLIMCAPS retrieves a total of 18 cloud parameters for each retrieval scene – 9 α 's for every C_{TP} layer – from a subset of channels sensitive to clouds in the troposphere. CLIMCAPS does not retrieve any cloud microphysical properties.

2. Does CLIMCAPS retrieve soundings from clear-sky scenes only?

No. CLIMCAPS retrieves atmospheric profile variables from clear and partly cloudy scenes. In fact, CLIMCAPS retrievals pass quality control in scenes with as much as 80-90% cloud cover. This said, it is important to distinguish that CLIMCAPS does not retrieve sounding profiles *through* cloud fields, but instead uses a technique known as ‘cloud clearing’ to remove the radiative effects of clouds from the infrared measurements. One can understand CLIMCAPS retrievals as representing the clear portion of the atmosphere past or around cloud fields in the target scene.

3. What is cloud clearing?

Cloud clearing is a linear extrapolation technique that aggregates spectral channels from 9 FOVs with varying degrees of cloud cover into a single set of channels that represents the clear state of the atmosphere around clouds in the target scene (or FOR).

Cloud clearing requires no prior knowledge of clouds in a target scene, nor does it depend on radiative transfer calculations through clouds. It is a simple technique that uses the spatial variability in cloud cover among 9 FOVs as information content to linearly derive a set of cloud-free infrared channels. If there is no variability in cloud cover among the 9 FOVs in a FOR, then their spatial information content is zero and cloud clearing fails. In turn, the higher the cloud heterogeneity among a cluster of FOVs, the higher the spatial information content and more accurate the cloud clearing.

4. Does cloud clearing impact measurement information content?

Yes, but only for the infrared channels. Elsewhere (e.g., Smith and Barnett, 2020) we describe how CLIMCAPS information content is a function of instrument noise, measurement and scene-specific uncertainty, channel selection and channel weighting functions. Cloud clearing impacts CLIMCAPS information content by affecting the random instrument noise (NEN) and scene-specific uncertainty. We quantify this as **ampl_eta** and **etarej** (Section 2), propagate them into subsequent retrievals as described in Smith and Barnett (2019) and report them in the CLIMCAPS product file as diagnostic metrics for use during analysis.

Section 2: Brief summary of methods

Cloud clearing uses the spatial heterogeneity in cloud cover from an array of 3 x 3 FOVs as information content to derive a set of infrared spectral channels that represent the cloud-cleared

(or cloud-free) atmosphere at a target scene (FOR). This technique was first developed by (Chahine, 1977, 1982; Smith, 1968) and later adopted by the AIRS Science Team for use in the AIRS retrieval systems (Susskind et al., 2003, 2014, 2017) as well as CLIMCAPS (Smith and Barnett, 2019, 2020). This technique is well documented in peer review publications and reports (link to AIRS V6 user guides). Our goal here is to give a brief summary.

The main purpose of cloud clearing is to improve retrieval yield and allow observation of the vertical atmospheric state in partly cloudy conditions. CLIMCAPS is a global product and cloud clearing allows the successful retrieval of sounding variables from ~75% of all IR+MW measurements made in one day.

The purpose of CLIMCAPS cloud parameter retrievals (cloud fraction and cloud top pressure) is to determine which channels to ‘cloud clear’ and which to simply average. We write the retrieved cloud parameters to the Level 2 product file to represent the full atmospheric state and enhance diagnostic and data filtering capability in subsequent analyses.

We gave a graphic depiction of the CLIMCAPS retrieval flow elsewhere {insert link} and elaborate here on those steps that mention cloud top pressure (C_{TP}), cloud fraction (α) and cloud clearing (CC).

- Define the a-priori for two cloud layers within each CLIMCAPS retrieval footprint (FOR) as follows: $[\alpha_1 = 0.5, C_{TP1} = 350]$, $[\alpha_2 = 0.25, C_{TP2} = 800]$.
- Retrieve $[\alpha_1, C_{TP1}]$, $[\alpha_2, C_{TP2}]$ for each FOR from a subset of infrared channels using the a-priori variables as defined for temperature ($T(p)$), water vapor (H_2O), trace gases and surface variables.
- Based on the values of α and C_{TP} , determine which channels to ‘cloud clear’ and which to simply average.
- Once all atmospheric state variables have been retrieved, retrieve α for each FOV, C_{TP} for each FOR and derive cloud cleared radiance channels for each FOR.

Notes:

- Unlike AIRS V6, CLIMCAPS does not iterate cloud clearing and performs it only once before retrieving the atmospheric state variables.
- Cloud fraction and cloud top pressure have the same radiative effect in top-of-atmosphere infrared measurements. This makes them difficult to retrieve. We constrain our solution to retrieving only two layers of C_{TP} for the entire CLIMCAPS footprint (FOR), $9 \times \alpha$ for each FOV.
- We use the same channel set to retrieve both C_{TP} and α , which is a subset of the channels used in cloud clearing.
- Channels in the infrared window region are always cloud cleared, even if their cloud fraction retrievals are close to or equal to zero. This is a precautionary measure since space-based infrared measurements have, by definition, reduced observing capability in the boundary layer.

- All other channels are sometimes cloud cleared, sometimes not, depending on the scene-specific CTP and α values.
- Cloud clearing is more accurate the higher the cloud contrast among FOVs in a retrieval footprint (Figure 48).

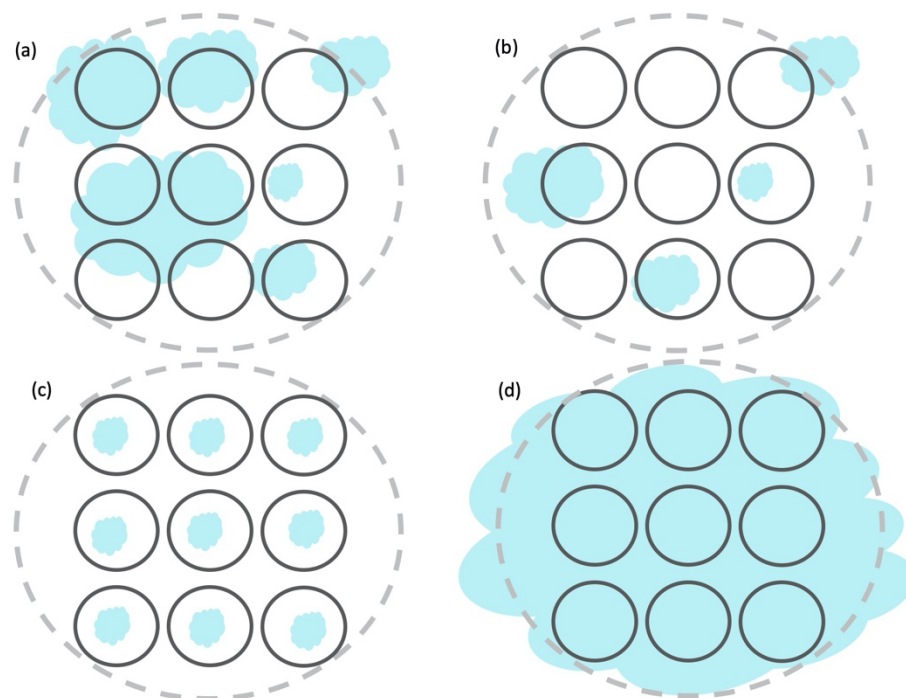


Figure 48: The CLIMCAPS retrieval footprint is the field-of-regard (FOR; grey dashed circles) that consists of 3 x 3 instrument fields-of-view (FOV; black solid circles). CLIMCAPS aggregates the 9 FOVs into a single spectrum from which it then retrieves a set of atmospheric profile variables. Cloud fraction is the only variable that CLIMCAPS retrieves for each FOV (9 per FOR); all other retrieval variables represent conditions within the FOR (~50km at nadir, ~150km at edge-of-scan). Here we illustrate four typical cloudy scenarios encountered by CLIMCAPS: (a) partly cloudy FOR where all FOVs have cloud fraction > 0.0 (i.e., not clear) but no two FOVs has the same cloud fraction, (b) partly cloudy FOR where some FOVs have no clouds (cloud fraction = 0), (c) partly cloudy where each FOV has the exact same cloud fraction (no contrast), and (d) overcast FOR where all FOVs have cloud fraction = 1.0. Cloud clearing is accurate (i.e., small brightness temperature residuals with low **etarej** values) in (top row) spatial *heterogeneous* retrieval footprints, but fails (i.e., large brightness temperature residuals with high **etarej** values) in (bottom row) spatial *homogeneous* scenes.

Readers can refer to studies that compared AIRS cloud retrievals to a host of other observations (Kahn et al., 2014, 2015; Nasiri et al., 2011; Weisz et al., 2007; Wong et al., 2015, 2015; Wu et al., 2009; Yue et al., 2011). There are also studies that demonstrate the value of AIRS cloud cleared radiances in data assimilation (Reale et al., 2008, 2018).

1. Relevant CLIMCAPS cloud and cloud clearing product fields

Note that we refer, here, to the CLIMCAPS Level 2 retrieval product. The cloud cleared radiances are written to and distributed as a separate product not discussed here.

Within the netCDF files, we highlight a few fields that are relevant to clouds. The fields have dimensions that correspond to the following variables:

- `atrack` = 30 (number of retrieval footprints along an instrument scanline)
- `xtrack` = 45 (number of scanlines grouped together in a CLIMCAPS file)
- `fov` = 9 (number of fields of view in a CLIMCAPS footprint)
- `cld_lay` = 2 (number of cloud retrieval layers)
- `air_pres_lay` = 100 (number of profile retrieval layers)

– Retrieved variables

- **`cld_frac(atrack, xtrack, fov, cld_lay)`**: cloud fraction retrievals for each field-of-view (AIRS or CrIS instrument footprint) and up to two cloud layers from a subset of infrared channels.
- **`aux/for_cld_top_pres_2lay(atrack, xtrack, cld_lay)`**: cloud top pressure retrievals for up to two layers of clouds on each CLIMCAPS footprint (or field-of-regard) from a subset of infrared channels.
- **`mw_cld_phase(atrack, xtrack, air_pres_lay)`**: cloud ice detection flag for every retrieval layer using information in microwave channels; 0 means the CLIMCAPS footprint at a target layer has only liquid clouds or is cloud free, while 1 means that ice clouds were detected.

– Derived variables

The methods used in deriving these cloud variables are the same as those used in the AIRS retrieval system for AIRS/AMSU and CrIS/ATMS. We refer the reader to Susskind et al. (2017) for a full description.

- **`cld_top_pres(atrack, xtrack, fov, cld_lay)`**: this is the **`for_cld_top_pres_2lay`** retrieval but reported on every field-of-view (FOV) to allow easy match-ups with the **`cld_frac`** field.
- **`cld_top_temp(atrac, xtrac, fov, cld_lay)`**: this is the value from **`air_temp`** (retrieved temperature profile) that corresponds to the cloud top pressure retrieval (**`for_cld_top_pres_2lay`**). Even though this field is reported on each FOV, it represents the temperature at the FOR (CLIMCAPS retrieval footprint) scale. The lack of variation in `cld_top_temp` across the FOVs as reported in this field does, therefore, not mean a real lack of variation across the FOR.
- **`num_cld(atrack, xtrack, fov)`**: number of cloud layers with nonzero cloud fraction as depicted by **`cld_frac`** for each FOV.
- **`aux/for_cld_frac_tot(atrack, xtrack)`**: cloud fraction across all cloud layers and FOVs to represent the total cloud cover for the target retrieval scene.
- **`aux/for_cld_top_pres_tot(atrack, xtrack)`**: cloud top pressure for the retrieval scene (FOR) as the total for all cloud layers and FOVs. It is calculated as the weighted sum of

the cloud top pressure from both cloud layers, divided by the sum of the cloud fraction from both cloud layers as follows:

$$\text{ctp_wght} = (\text{cld_top_pres}(1) * \text{cld_frac}(1)) + (\text{cld_top_pres}(2) * \text{cld_frac}(2))$$

$$\text{for_cld_top_pres_tot} = \text{ctp_wght} / (\text{cld_frac}(1) + \text{cld_frac}(2))$$

- **aux/for_cld_frac_2lay(atrack, xtrack, cld_lay)**: total cloud fraction across all FOVs for two layers. This is similar to **for_cld_frac_tot**.
- **cldfrac_500(atrack, xtrack)**: the total cloud fraction of all clouds below 500 hPa over the retrieval footprint (FOR). This is similar to **for_cld_frac_tot** but only for those clouds in the lower troposphere.

– *Uncertainty metrics*

- **aux/etarej(atrack, xtrack)**: The cloud clearing radiance error in brightness temperature units [Kelvin] calculated as the difference between a simulated clear-sky spectrum and the derived cloud cleared spectrum at the target retrieval scene. **Etarej** quantifies the quality of cloud clearing by indicating how well the cloud-cleared radiance represents the clear-sky state around the clouds at that scene. Smaller values of **etarej** indicate successful cloud clearing and a high confidence in the removal of clouds from the infrared radiance measurements. Higher values of **etarej** indicate a lower confidence in cloud clearing and retrievals should be interpreted as being ‘contaminated’ by residual undetected clouds.
- **aux/ampl_eta(atrack, xtrack)**: The amplification factor (**ampl_eta**) quantifies how much the random instrument noise (NEN or NEdT in units Kelvin) was amplified ($\text{ampl_eta}(i,j) > 1$) or damped ($\text{ampl_eta}(i,j) < 1$) as a result cloud clearing.
- **aux/aeff_end(atrack, xtrack)**: The effective amplification factor (**aeff_end**) is a compound metric that combines **random** instrument noise as scaled by the **ampl_eta** with **systematic** uncertainty due to spectral correlation.

References

- Chahine, M. T.: Remote sounding of cloudy parameters II. Multiple cloud formations, J. Atmos. Sci., 34, 744–757, 1977.
- Chahine, M. T.: Remote sensing of cloud parameters, J. Atmos. Sci., 39, 159–170, 1982.
- Kahn, B. H., Irion, F. W., Dang, V. T., Manning, E. M., Nasiri, S. L., Naud, C. M., Blaisdell, J. M., Schreier, M. M., Yue, Q., Bowman, K. W., Fetzer, E. J., Hulley, G. C., Liou, K. N., Lubin, D., Ou, S. C., Susskind, J., Takano, Y., Tian, B. and Worden, J. R.: The Atmospheric Infrared Sounder version 6 cloud products, Atmospheric Chemistry and Physics, 14(1), 399–426, doi:10.5194/acp-14-399-2014, 2014.
- Kahn, B. H., Schreier, M. M., Yue, Q., Fetzer, E. J., Irion, F. W., Platnick, S., Wang, C., Nasiri, S. L. and L’Ecuyer, T. S.: Pixel-scale assessment and uncertainty analysis of AIRS and MODIS ice cloud optical thickness and effective radius: AIRS AND MODIS ICE CLOUDS, Journal of Geophysical Research: Atmospheres, 120(22), 11,669–11,689, doi:10.1002/2015JD023950, 2015.
- Nasiri, S. L., Van T. Dang, H., Kahn, B. H., Fetzer, E. J., Manning, E. M., Schreier, M. M. and Frey, R. A.: Comparing MODIS and AIRS Infrared-Based Cloud Retrievals, Journal of

- Applied Meteorology and Climatology, 50(5), 1057–1072, doi:10.1175/2010JAMC2603.1, 2011.
- Reale, O., Susskind, J., Rosenberg, R., Brin, E., Liu, E., Riishojgaard, L. P., Terry, J. and Jusem, J. C.: Improving forecast skill by assimilation of quality-controlled AIRS temperature retrievals under partially cloudy conditions, *Geophysical Research Letters*, 35(8), doi:10.1029/2007GL033002, 2008.
- Reale, O., McGrath-Spangler, E. L., McCarty, W., Holdaway, D. and Gelaro, R.: Impact of Adaptively Thinned AIRS Cloud-Cleared Radiances on Tropical Cyclone Representation in a Global Data Assimilation and Forecast System, *Weather and Forecasting*, 33(4), 909–931, doi:10.1175/WAF-D-17-0175.1, 2018.
- Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), *Remote Sensing*, 11(10), 1227, doi:10.3390/rs11101227, 2019.
- Smith, N. and Barnet, C. D.: CLIMCAPS Observing Capability for Temperature, Moisture and Trace Gases from AIRS/AMSU and CrIS/ATMS, preprint, *Gases/Remote Sensing/Data Processing and Information Retrieval*, 2020.
- Smith, W. L.: An improved method for calculating tropospheric temperature and moisture from satellite radiometer measurements, *Monthly Weather Review*, 96(6), 387–396, doi:10.1175/1520-0493(1968)096<0387:AIMFCT>2.0.CO;2, 1968.
- Susskind, J., Barnet, C. D. and Blaisdell, J. M.: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE TGRS*, 41, 390–409, 2003.
- Susskind, J., Blaisdell, J. M. and Iredell, L.: Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm, *Journal of Applied Remote Sensing*, 8(1), 084994, doi:10.1117/1.JRS.8.084994, 2014.
- Susskind, J., Blaisdell, J., Kouvaris, L. and Iredell, L.: CrIS CHART Retrieval Algorithm ATBD, Algorithm Theoretical Basis Document, NASA GSFC. [online] Available from: https://docsserver.gesdisc.eosdis.nasa.gov/public/project/SNPP/SNPP_limited_edition/SNPP_CrIMSS.CHART_V1.ATBD.pdf, 2017.
- Weisz, E., Li, J., Menzel, W. P., Heidinger, A. K., Kahn, B. H. and Liu, C.-Y.: Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals, *Geophysical Research Letters*, 34(17), doi:10.1029/2007GL030676, 2007.
- Wong, S., Fetzer, E. J., Schreier, M., Manion, G., Fishbein, E. F., Kahn, B. H., Yue, Q. and Irion, F. W.: Cloud-induced uncertainties in AIRS and ECMWF temperature and specific humidity: Cloud-dependent AIRS V6 validation, *Journal of Geophysical Research: Atmospheres*, 120(5), 1880–1901, doi:10.1002/2014JD022440, 2015.
- Wu, D. L., Ackerman, S. A., Davies, R., Diner, D. J., Garay, M. J., Kahn, B. H., Maddux, B. C., Moroney, C. M., Stephens, G. L., Veefkind, J. P. and Vaughan, M. A.: Vertical distributions and relationships of cloud occurrence frequency as observed by MISR, AIRS, MODIS, OMI, CALIPSO, and CloudSat, *Geophysical Research Letters*, 36(9), doi:10.1029/2009GL037464, 2009.

Yue, Q., Kahn, B. H., Fetzer, E. J. and Teixeira, J.: Relationship between marine boundary layer clouds and lower tropospheric stability observed by AIRS, CloudSat, and CALIOP, *Journal of Geophysical Research*, 116(D18), doi:10.1029/2011JD016136, 2011.

Chapter 5: Working with CLIMCAPS Averaging Kernels

Averaging kernels (AKs) quantify the information content of Optimal Estimation (OE) retrieval systems and are key in understanding retrieval uncertainty and error. CLIMCAPS outputs a square matrix of AKs for each retrieval variable at every scene (3 x 3 fields of view) to allow diagnostic evaluation. We explain in detail how CLIMCAPS averaging kernels are calculated in Smith and Barnett (2020), and contrast our method with that originally proposed by (Rodgers, 2000).

Averaging kernels are used in data inter-comparison studies (Rodgers and Connor, 2003; Maddy and Barnett, 2008; Maddy et al., 2009; Gaudel et al., 2018; Iturbide-Sanchez et al., 2017; Smith and Barnett, 2020) and data assimilation models (Levelt et al., 1998; Clerbaux et al., 2001; Yudin, 2004; Segers et al., 2005; Pierce et al., 2009; Liu et al., 2012).

Section 1: Interpreting CLIMCAPS Averaging Kernels

Eric Maddy gave a presentation in 2006 at an AIRS Science Team Meeting specifically focused on AKs.

E. Maddy et al. 2006: “**Application of AIRS V5.0 Averaging Kernels**”. AIRS Science Team Meeting, 26-29 Sep 2006, Greenbelt, MD,
https://airs.jpl.nasa.gov/documents/science_team_meeting_archive/2006_09/slides/maddy_0609_airsstm.pdf

We refer the reader to his presentation because it provides an excellent overview of AKs in general and CLIMCAPS AKs specifically. CLIMCAPS Version 2.0 (V2) is an AIRS Science Team heritage algorithm that is built on the AIRS Version 5.0 method to ensure robust retrievals across the globe and across all seasons, day and night.

What are averaging kernels?

- Averaging kernels are a linear representation of the vertical weighting of retrievals.
 - Related to the amount of information determined from the radiances and how much is due to the first guess [Rodgers, 1976].
 - To some degree avoids aliasing comparisons of in situ measurements vs. retrievals due to incorrect first guesses.
 - Enables assessment of where vertically we have information.
 - Related to the vertical resolution of retrievals [Backus and Gilbert, 1969; Conrath, 1972; Rodgers, 1976; Purser and Huang, 1993]
 - Required by modelers to properly use AIRS trace gas products.
 - Enables assessment of retrieval skill on a case by case basis.
- In the IDEAL case (no damping): $A = I$: the identity matrix

Eric Maddy's seminal paper (Maddy and Barnett, 2008) on averaging kernel application forms the theoretical background to discussions here. Nalli et al. (2013) give a good overview of the Maddy paper with a focus on data validation methods. Iturbide-Sanchez et al. (2017), in turn, discuss how AKs can be interpreted as a metric of vertical resolution. Our paper on CLIMCAPS averaging kernels highlight their value in diagnostic retrieval evaluation (Smith and Barnett 2020).

The CLIMCAPS product file contains symmetric AK matrices for every retrieval of temperature (air_temp), water vapor (H₂O_vap), ozone (O₃), methane (CH₄), carbon dioxide (CO₂), carbon monoxide (CO) and nitric acid (HNO₃) irrespective of whether the retrieval passed QC or not. These are reported in the **ave_kern/** subgroup with field names suffixed by ***_ave_kern**. The CLIMCAPS V2 product file does not contain AKs for nitrous oxide (N₂O) and Sulphur dioxide (SO₂) because we are unaware of any user applications that require them. This may change in future.

We published different versions of Figures 51 through 54 in Smith and Barnett (2020) with full explanation and discussion. Here, we supplement the discussion in that paper with figures of CLIMCAPS AKs for different seasons and scenes for the sake of a complete representation of CLIMCAPS information content from AIRS/AMSU on Aqua and CrIS/ATMS on JPSS-1.

Figures 49 and 50 depict the individual AKs for CLIMCAPS air_temp and H₂O_vap retrievals at specific scenes, while Figures 51 through 54 depict the mean of averaging kernel matrix diagonal vectors for different latitudinal zones on 1 July 2018. The error bars in these figures represent the standard deviation of the diagonal vectors across the latitude zone. In Figures 55 and 56 we plot the AK diagonal vectors at ~500 hPa for CLIMCAPS H₂O_vap and ~700hPa for CLIMCAPS air_temp at every retrieval scene to give a global view of information content variability from ascending orbits of CLIMCAPS-Aqua and CLIMCAPS-JPSS1 on 15 December 2018. We did not apply any quality control filtering to the AK profiles or maps because AKs are not dependent on whether a retrieval succeeds or fails. Instead, averaging kernels depict the potential CLIMCAPS has for retrieving a target variable at each scene. Scenes with much lower (higher) AK values can be interpreted as CLIMCAPS having low (high) information content, and the retrievals probably failed (succeeded).

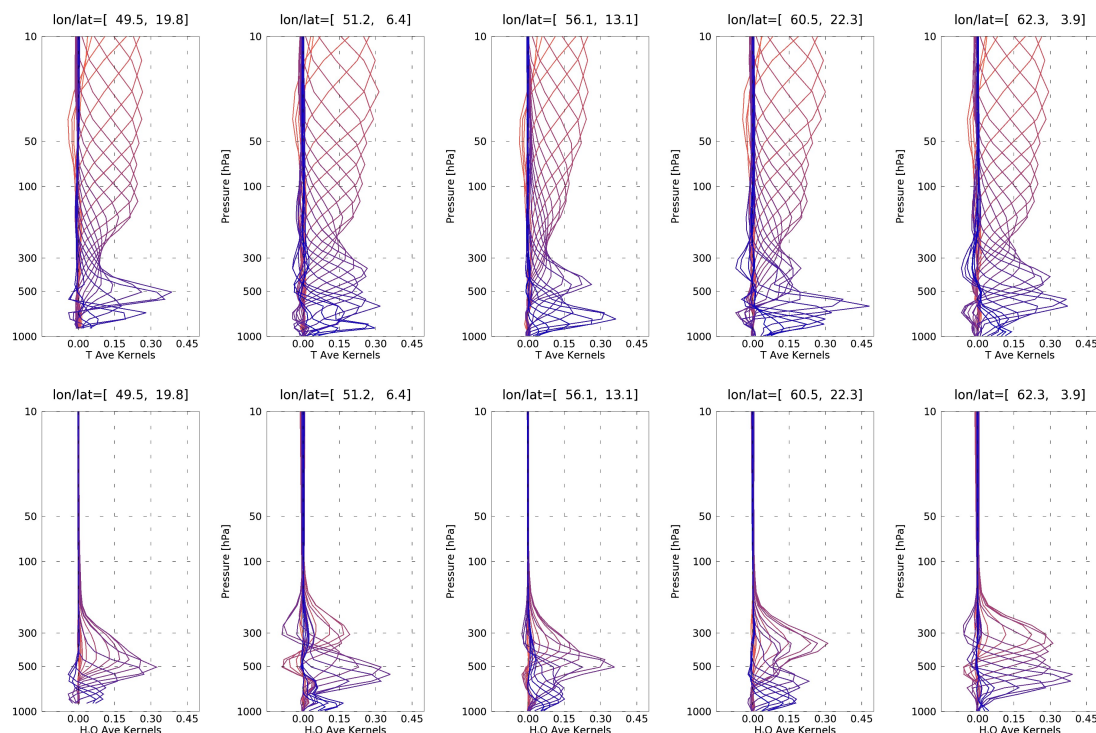


Figure 49: Scene-dependence of CLIMCAPS-Aqua averaging kernels for coincident (top row) temperature (air_temp) and (bottom row) water vapor (h2o_vap) retrievals at five scenes (left to right) on 1 July 2018 from Granule 60 with 13h30 local overpass time. The latitude/longitude coordinates are listed at the top of each figure. Averaging kernels quantify and characterize the signal-to-noise ratio of an observing system and are affected by the scene-dependent effects (e.g., temperature lapse rate, amount of gas molecules, surface emissivity and cloud uncertainty) as much as the measurement characteristics (e.g., spectral resolution, instrument calibration and noise). CLIMCAPS retrieves air_temp and h2o_vap sequentially each with a unique subset of channels, which means that the variation in these averaging kernels are independent of each other.

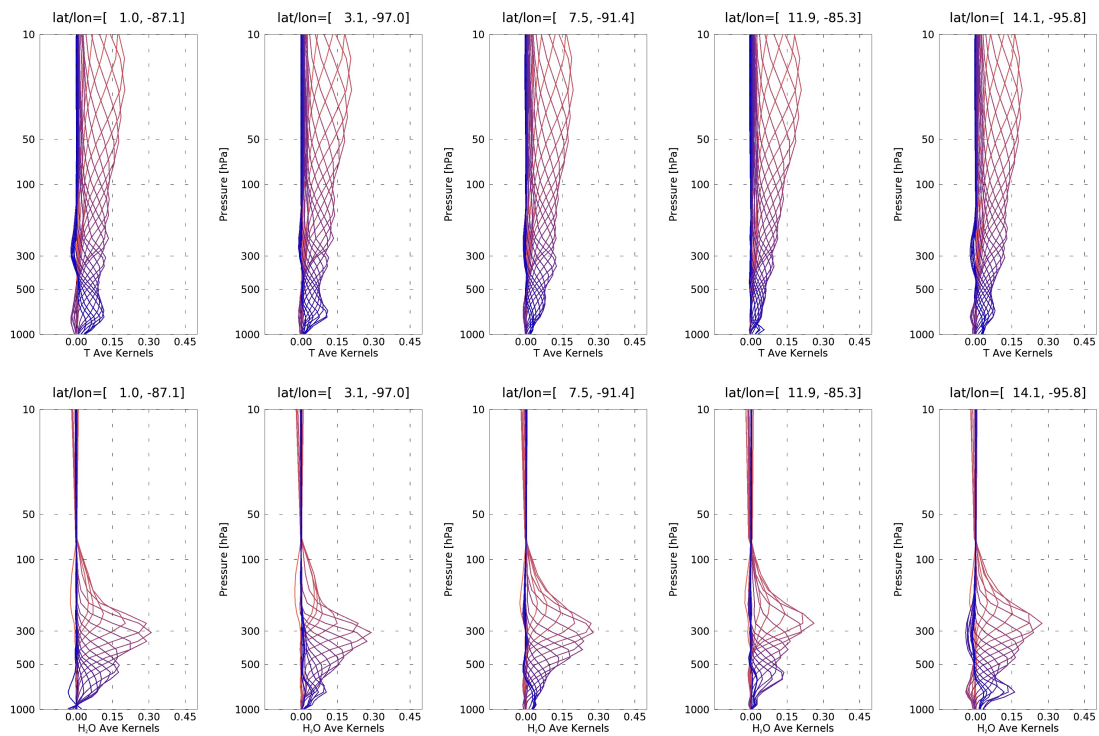


Figure 50: Same as Figure 48 but for CLIMCAPS-JPSS1 averaging kernels from Granule 97 on 1 July 2018 and 13h30 local overpass time.

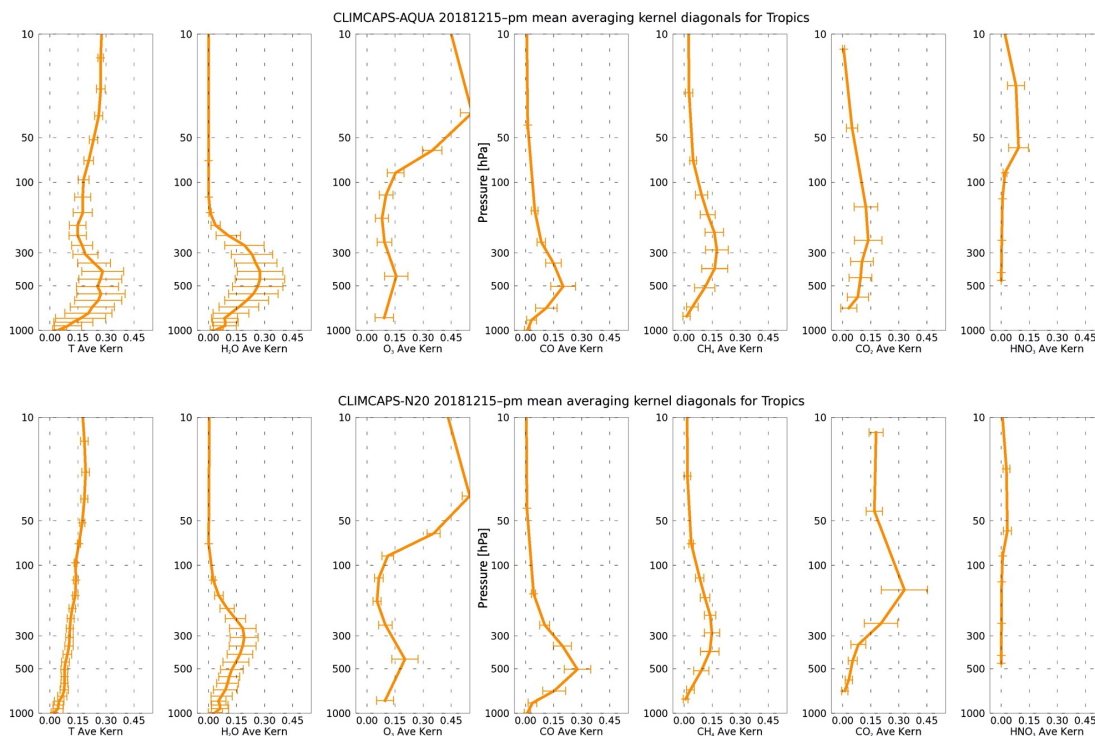


Figure 51: The mean (solid line) and standard deviation (error bars) of averaging kernel matrix diagonal vectors in the tropics (30°S to 30°N) on 1 July 2018 and from (top) CLIMCAPS-Aqua and (bottom) CLIMCAPS-JPSS1, both ascending orbits. The error bars indicate the degree to which the averaging kernel diagonal vectors vary spatially across the latitudinal zonal. CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnet 2020).

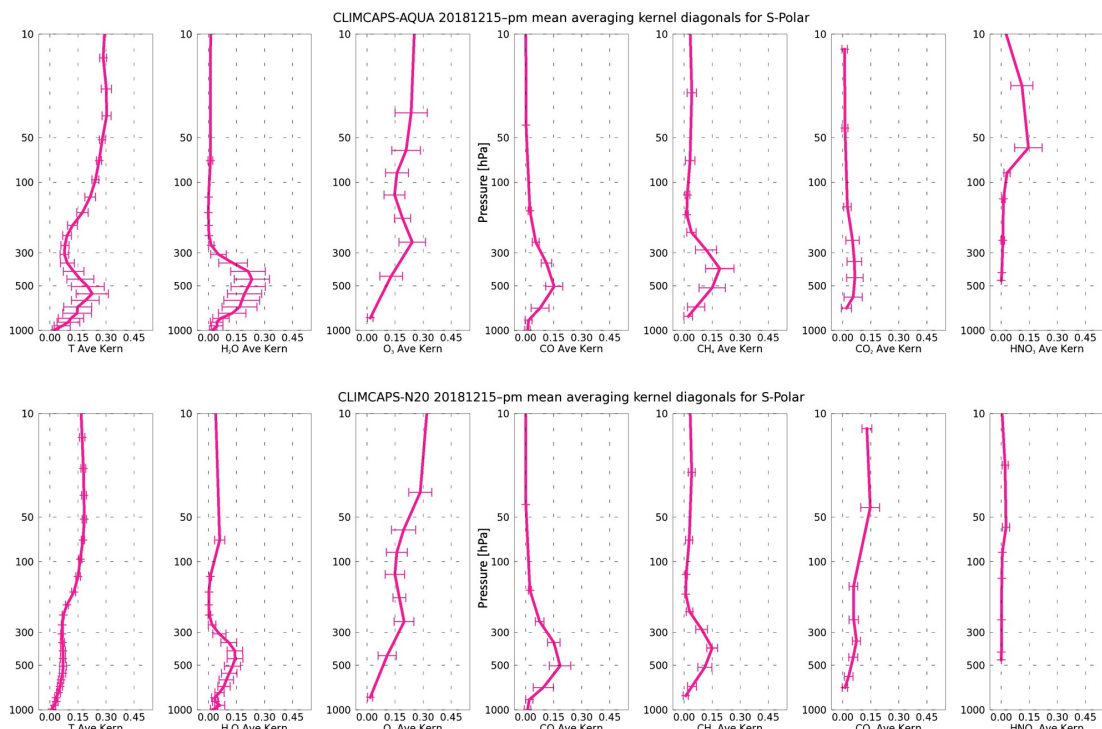


Figure 52: Same as Figure 50 but for southern polar zone (90°S to 60°S). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020).

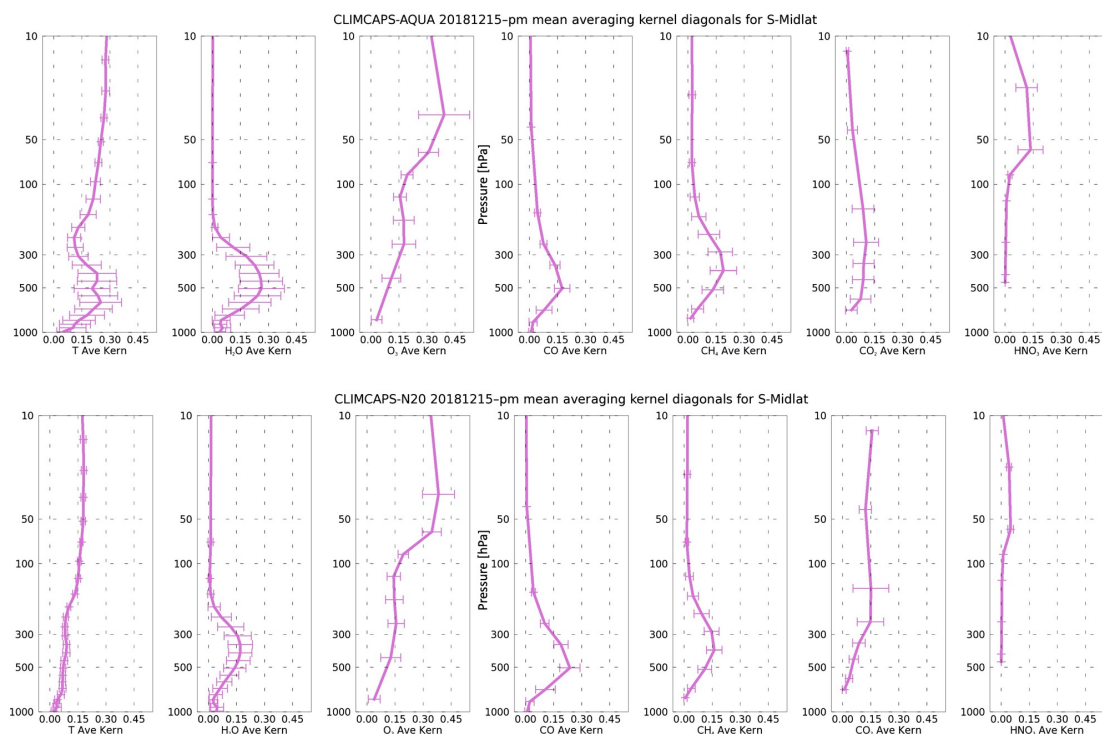


Figure 53: Same as Figure 50 but for southern mid-latitude zone (60°S to 30°S). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020).

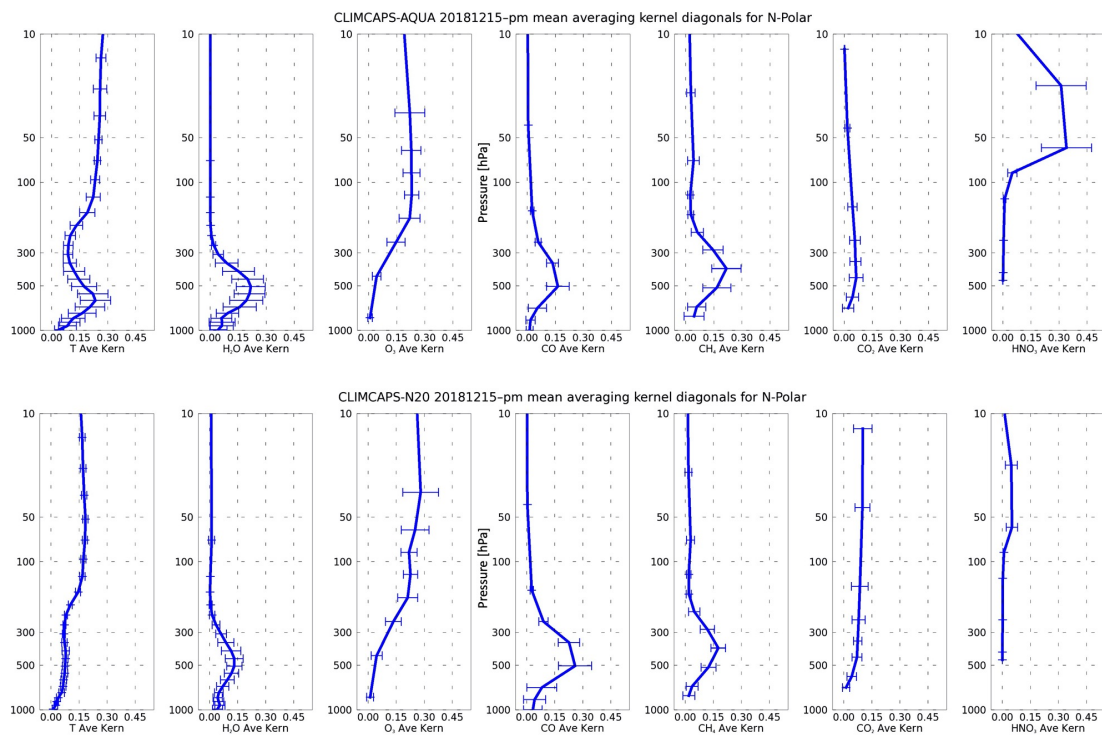


Figure 54: Same as Figure 50 but for northern polar zone (60°N to 90°N). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020).

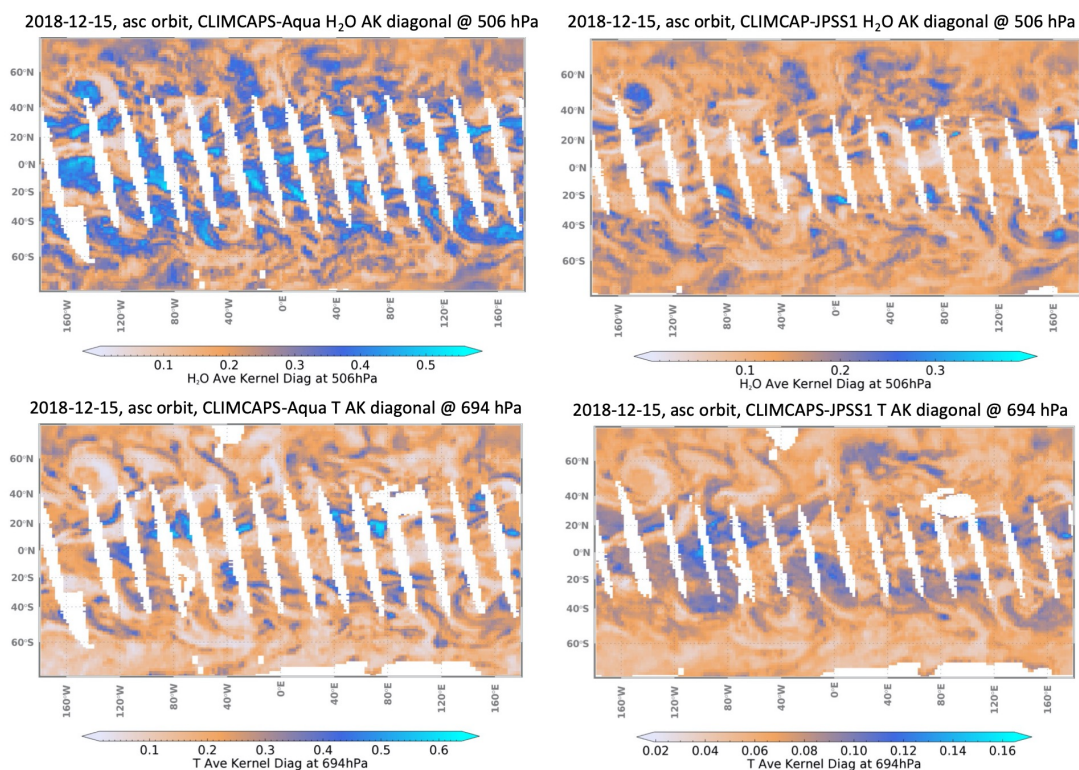


Figure 55: Spatial variation of CLIMCAPS information content for different retrieval variables and different space-based instruments from their ascending orbits (13h30 local overpass time) on 15 December 2018. (Top row) A global view of CLIMCAPS averaging kernel values at ~500 hPa for H₂O vapor retrievals from (top left) AIRS/AMSU on Aqua, and (top right) CrIS/ATMS on JPSS-1. (Bottom row) A global view of CLIMCAPS averaging kernel values at ~700 hPa for temperature retrievals from (bottom left) AIRS/AMSU on Aqua, and (bottom right) CrIS/ATMS on JPSS-1.

The spatial patterns of the CLIMCAPS AKs in Figures 55 and 56 appear to be influenced by cloud fields. CLIMCAPS does not depend on knowledge of clouds to do retrievals, i.e., it does not use radiative transfer calculations through clouds. Instead, it performs cloud clearing to remove the radiative effect of clouds from the instrument measurements. We can, however, attribute the cloud patterns observed here to the fact that CLIMCAPS calculates and propagates the radiance uncertainty due to cloud clearing into the measurement error covariance matrix, which is one of the key parameters in calculating AKs (Smith and Barnett, 2019, 2020).

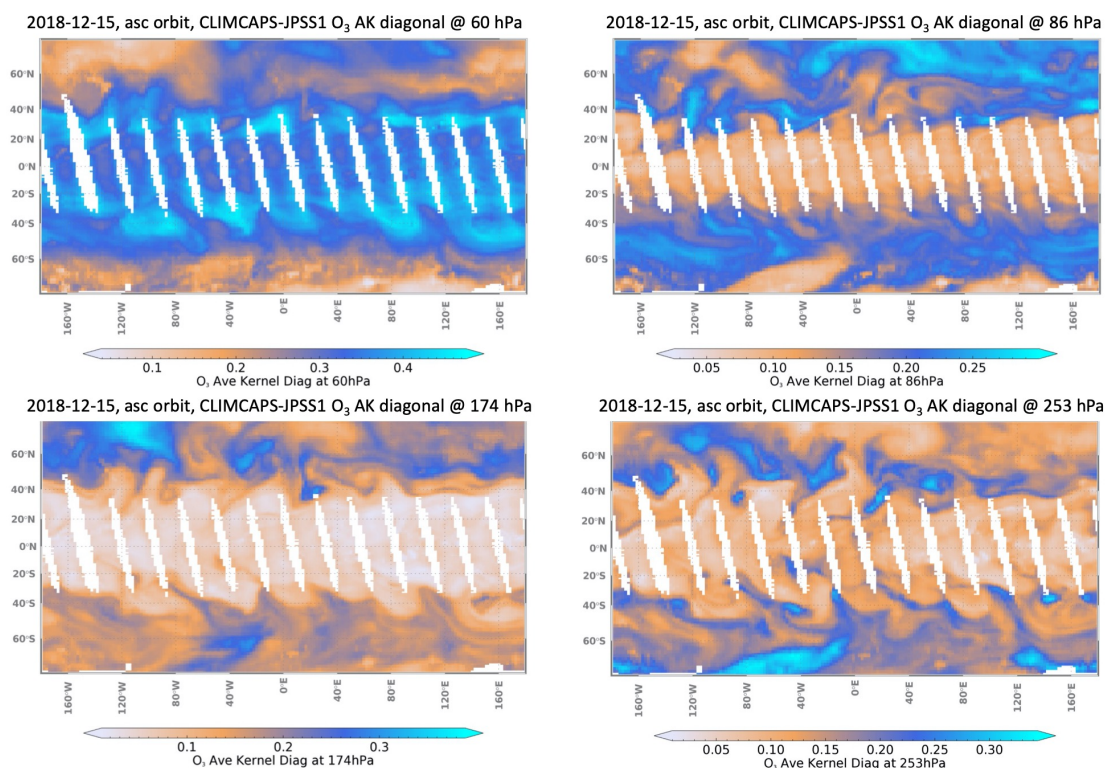


Figure 56: Spatial and vertical variation of CLIMCAPS-JPSS1 ascending orbit (13h30 local overpass time) information content for ozone at different pressure values from the stratosphere to upper troposphere on 15 December 2018. The panels are CLIMCAPS-JPSS1 ozone averaging kernel diagonal values at (top left) 60 hPa (top right) 86 hPa, (bottom left) 174 hPa, and (bottom right) 253 hPa.

Section 2: Trapezoid State Functions

CLIMCAPS calculates Jacobians (weighting functions), not on the standard 100 retrieval pressure layers (`air_pres_lay`), but on a reduced set of overlapping pressure layers, referred to as trapezoid state functions. The use of trapezoid state functions was first developed by (Susskind et al., 2003) and remains a key feature of CLIMCAPS (Smith and Barnett, 2019, 2020) and AIRS V6 (Susskind et al., 2014) today.

The set of trapezoid state functions is unique to each retrieval variable. In Table 5 we list the relevant CLIMCAPS product fields, the number of state functions per variable as well as the upper and lower pressure layer values (i.e., the mid-point of the trapezoid ‘face’). Variables for which the radiance measurements have a higher (lower) information content, have more (fewer) state functions. The number of trapezoid state functions define the number of Jacobians and thus the dimension of the AK matrix for each retrieval variable. The full set of trapezoid state functions are visually depicted for seven retrieval variables in Figure 56.

– Value of trapezoid state functions and how do they work in practice

Trapezoids greatly speed up Optimal Estimation retrievals with fewer calls to the forward model. All forward model calculations in CLIMCAPS are performed on the standard 100 pressure

layers, but the Jacobians are calculated on the trapezoid state functions as brute force perturbation of the a-priori profiles (See Eq. 2 in Smith and Barnett 2020). Each a-priori profile is perturbed on multiple pressure layers at once according to the shape and magnitude of a trapezoid state function. For air_temp, CLIMCAPS calls the Stand-alone AIRS Radiative Transfer Algorithm (SARTA; Strow et al., 2003) 30 times, not 100 times, and for O₃, only nine times. Moreover, storing AK matrices on trapezoid pressure layers instead of the 100 retrieval layers significantly reduces CLIMCAPS product size. This is especially important for a global product across decades.

The AIRS science team determined the shape and distribution of trapezoids empirically to more closely represent the real information content of the measurements. Even though retrievals are reported on 100 levels (or layers) the number of levels (layers) the measurements have information content for is much less.

Table 5: Summary of CLIMCAPS averaging kernel variables in the product file. All fields listed here are from the **ave_kern/** group inside the netCDF file. P_{top} is the pressure layer value [hPa] of the top-most Trapezoid state function, and P_{bot}, the pressure layer value of the bottom-most state function. We read P_{top} and P_{bot} from the corresponding ***_func_pres** fields.

Retrieval Variable	Averaging kernel matrix [ave_kern/]	Trapezoid state function hinge-points [ave_kern/]	Number of trapezoid state functions	P _{top} [hPa]	P _{bot} [hPa]
Air_temp	air_temp_ave_kern	air_temp_func_idxxs	30	0.08	1056.4
H ₂ O_vap	h2o_ave_kern	h2o_func_idxxs	21	5.5	1056.4
O ₃	o3_ave_kern	o3_func_idxxs	9	2.9	822.6
CH ₄	ch4_ave_kern	ch4_func_idxxs	11	1.7	943.4
CO	co_ave_kern	co_func_idxxs	9	1.5	999.3
CO ₂	co2_ave_kern	co2_func_idxxs	8	1.9	888.8
HNO ₃	hno3_ave_kern	hno3_func_idxxs	8	1.3	733.92

Each CLIMCAPS product file has a vector in the **ave_kern** netCDF group that defines the trapezoid pressure hinge-points (or indices) as the upper and lower limits of the trapezoid ‘face’. These are the ***_func_idxxs** fields (Table 1, column 3). Each variable has two additional fields that define their structure, namely ***_func_hbot**, and ***_func_htop**. A value of 1 indicate that the top trapezoid state function (i.e., the one that is closest to the top of atmosphere) is a trapezoid, whereas a value of 0 indicate that it is a wedge (Figure 57). Air_temp, CO₂ and HNO₃ all have ***_func_htop** = 0. CH₄, CO₂ and HNO₃ all have ***_func_hbot** = 0. These values become useful in the procedures we discuss below.

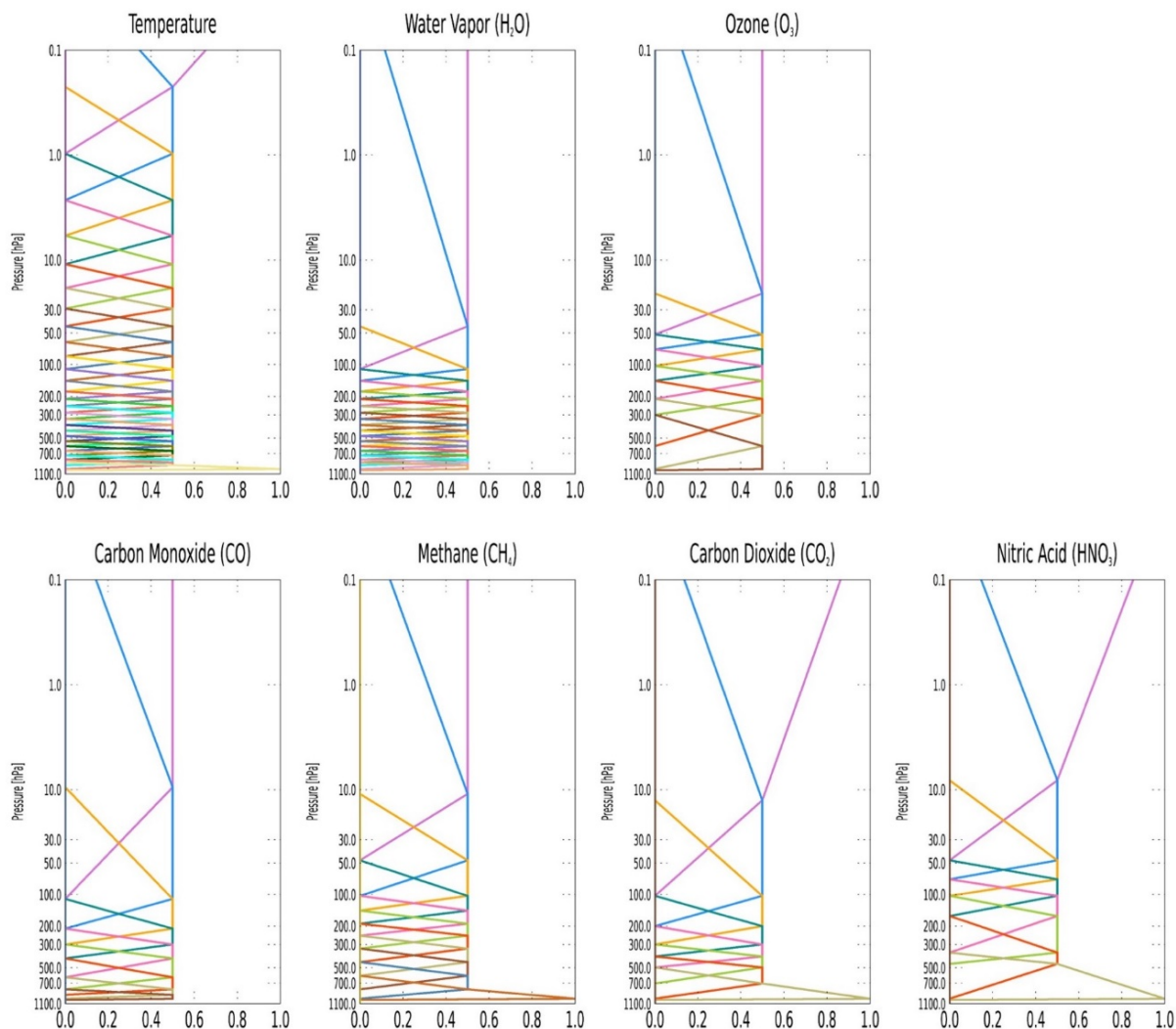


Figure 57: Trapezoid state functions for air_temp, h2o_vap, O₃, CO, CH₄, CO₂ and HNO₃, which define the pressure levels on which CLIMCAPS Jacobians (weighting functions) and thus their averaging kernels are calculated.

Note how the trapezoid state functions overlap. This is to mimic correlation across vertical layers, and avoid introducing spurious structure in the retrieved profiles.

Section 3: Convolution with a reference profile

We distinguish two scenarios where averaging kernels may have value in applications:

- Inter-comparison studies using data in their native formats, without the need to convert to a common framework. In such cases, CLIMCAPS AKs on their trapezoid layers as distributed in the product file is sufficient to characterize information content and diagnose retrieval uncertainty on a scene-to-scene basis. We give examples of how this can be done with CLIMCAPS AKs in Smith and Barnett (2020). Iturbide-Sanchez et al., (2017) give another example using NUCAPS AKs (a CLIMCAPS sister algorithm).

- Applications, such as data assimilation and validation, may require measurements to be defined within a common framework that includes transforming them to the same vertical grid and applying similar smoothing constraints. We refer to this data transformation as ‘data convolution’. A typical example is the validation of retrieved soundings using radiosondes that were convolved using the retrieval’s AKs as outlined in Maddy and Barnet (2008). In simple terms, two types of ‘convolutions’ exist; Eq. 14 shows an example where a reference profile is smoothed according to the retrieval’s averaging kernel, or vertical resolution; Eq. 15 is the similar to Eq. 14 but additionally combines the reference profile with the retrieval a-priori.

$$\tilde{x} = \mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+(x) \quad 14$$

$$\tilde{x} = x_a + \mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+(x - x_a) \quad 15$$

Where $x [j \times 1]$ is the reference profile (colloquially referred to as the ‘truth’ profile) from a different source, such as a radiosonde or reanalysis model profile interpolated to the standard 100 level retrieval grid, $\tilde{x} [j \times 1]$ is the solution or convolved reference profile, $x_a [j \times 1]$ the CLIMCAPS a-priori profile (e.g., MERRA2 interpolated in time, space and pressure to match CLIMCAPS retrieval scenes and vertical grid), $\mathbf{F} [L \times j]$ is the transformation matrix reconstructed from the hinge-points, $\mathbf{F}^+ [L \times j]$ is the pseudo inverse of \mathbf{F} , and $\mathbf{A} [L \times L]$ is the averaging kernel matrix on trapezoid layers. Index L indicates the coarse vertical grid as defined by the trapezoid state function layers above surface pressure and j indicates the standard 100 retrieval levels above surface pressure. For gases, the linearity of Eqs. 14 and 15 is achieved by using the log-form of the profiles, such that:

$$\tilde{x} = \text{Exp}[\log(x_a) + \mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+ \times (\log(x) - \log(x_a))] \quad 16$$

We can refer to the matrix $\mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+ [j \times j]$ as the effective averaging kernels of the 100-level retrieval profiles that describe the vertical correlation in the retrieval products, and thus their vertical resolution (Maddy and Barnet, 2008). If you wish to use CLIMCAPS averaging kernels on the 100 retrieval levels, then you need to calculate $\mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+$ following the instructions below.

The averaging kernel matrix, $\mathbf{A} [L \times L]$ is available for seven CLIMCAPS retrieval variables in the netCDF file as **ave_kern/*_ave_kern**. The a-priori profiles for three retrieval variables are available in the CLIMCAPS product file as **aux/fg_o3_mol_lay**, **aux/fg_h2o_vap_mol_lay** and **aux/fg_air_temp**. The a-priori profiles for CO, CO₂ and CH₄ should be calculated offline as detailed in the respective application guides. The only missing component is the calculation of the transformation matrix, \mathbf{F}_{Lj} (using the pressure hinge-points as recorded in the **ave_kern/*_func_indxs** product field) and its inverse, \mathbf{F}_{Lj}^+ as:

$$\mathbf{F}_{Lj}^+ = [\mathbf{F}_{Lj} \mathbf{F}_{Lj}^T]^{-1} \mathbf{F}_{Lj} \quad 17$$

We demonstrate how to calculate \mathbf{F} and \mathbf{F}^+ using pseudocode below.

1. What to expect when you convolve a reference profile

A common practice is to convolve radiosondes and ozonesondes to satellite sounding retrievals to ease interpretation in some applications. Sondes have much higher vertical resolution, vertical sampling and vertical structure, compared to satellite soundings. Inter-comparisons can become difficult without convolving the higher-resolution measurement to the lower-resolution one because differences may be due to variation in quality or may be more reflective of their fundamental measurement characteristics. We will illustrate this with an ozonesonde as example.

In Figure 58, left panel, there are five profiles; the ozone ‘sonde’ (solid black), the ‘retrieval’ (solid red) and the retrieval ‘first guess’, or a-priori (solid green). Applying Eq. 1 results in the ‘smoothed sonde’ (dash-dot-dash) and Eq. 2 in the ‘convolved sonde’ (dash-dash). We see here that the ‘sonde’ depict more vertical structure with a sharper tropopause transition. The ‘retrieval’, in contrast, has less vertical structure with a smoother transition from troposphere to stratosphere. The retrieval improved upon its ‘first guess’ (which clearly over-estimated the upper troposphere lower stratosphere, UTLS, O₃ concentrations) but not enough to align with the ‘sonde’. The ‘smoothed sonde’ has a smoother appearance, when compared to the original ‘sonde’ and the ‘convolved sonde’ aligns with the ‘retrieval’ because it is weighted by the ‘first guess’.

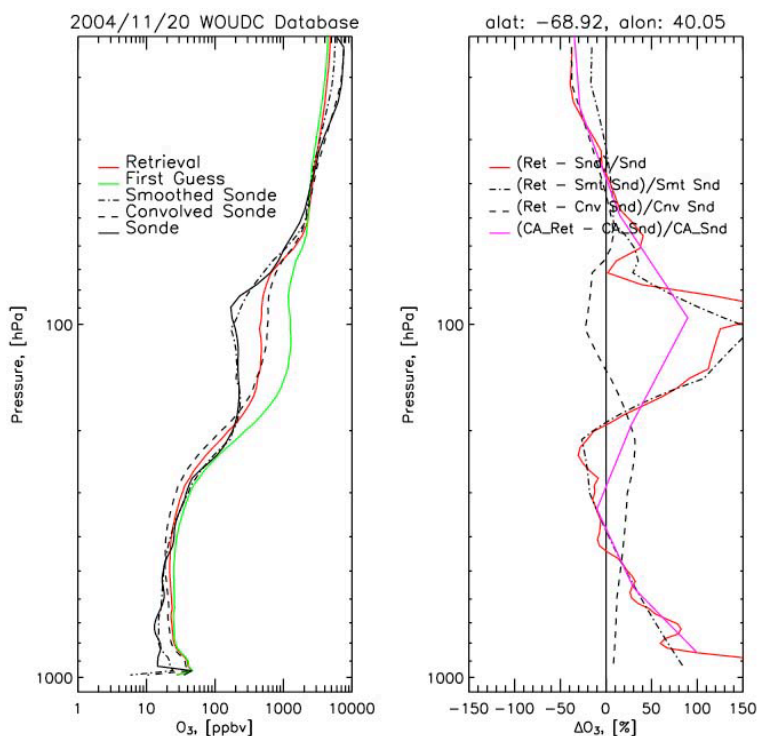


Figure 58: Demonstrating the value of convolving methods. Originally presented by Chris Barnett
in: http://www.weatherchaos.umd.edu/group_log/data/y0910/091019_weatherchaos_barnet.pdf

In Figure 58, right panel, we depict the percent bias and note how the convolved sonde bias as ‘(Ret – Cnv Snd)/Cnv Snd’ (dash-dash) oscillate around zero throughout the vertical column, and the smoothed sonde bias ‘(Ret – Smt Snd)/Smt Snd’ (dash-dot-dash) has high peaks in the lower

troposphere and UTLS. The convolved sonde statistics show us that the bias in the retrieval comes from the a-priori, not the radiance measurements or retrieval method.

2. Constructing the transformation matrix and its inverse

By ‘transformation matrix’, \mathbf{F} , we mean the reconstructed trapezoid state functions as an $[L \times j]$ matrix that transforms the averaging kernel matrix $[L \times L]$ from coarse layers (L , Table 1) to the standard 100 pressure grid (j) as follows: $\mathbf{F}_{Lj}^T \mathbf{A}_{LL} \mathbf{F}_{Lj}^+$ with dimension $[j \times j]$.

Four product fields are necessary to calculate the trapezoid state functions. They are the (1) 100-level pressure grid (**air_pres**), (2) shape of the top-most state function (**ave_kern/*_func_htop**), (3) shape of the bottom-most state function (**ave_kern/*_func_htop**), and (4) trapezoid state function indices (**ave_kern/*_func_indxs**). As shown in Table 1, each variable has a different number of indices. The trapezoid state function indices (also known as *hinge points*), range in value between 1 and 100, each representing a pressure level in **air_pres**. When you reconstruct the trapezoid state functions using these hinge-points, they will resemble those in Figure 56.

We refer the reader to two IDL routines – `slb2fin.pro` that is called by `calc_finv.pro`. These routines are available as a standalone IDL procedures in our github repository [insert link], and we also appended them to the end of this chapter. In Table 6 and its discussion below, we attempt to clarify `slb2fin.pro` and demonstrate how it creates a trapezoid state function using the hinge points [**ave_kern/*_func_indxs**] available in the CLIMCAPS product file.

Table 6: A summary of the output created by `slb2fin.pro` for a single trapezoid state function. Here we use CLIMCAPS O₃ as example and list the content of [**ave_kern/o3_func_indxs**] in column 2. The standard 100-level pressure array [**air_pres**] in hPa units is listed in column 3. All other column titles refer to the variable names as used in `slb2fin.pro` and `calc_finv.pro`. Column 4 (**func_ampl**) is defined in `calc_finv.pro` and used as one of the input variables to `slb2fin.pro`, and indicates that the fourth O₃ trapezoid state function should be calculated (row 4 of column 4 = 1.0). The last column (**slope**) defines the shape of the state function.

Row num	Input parameters									
	o3_func_indxs	air_pres [hPa]	func_ampl	IDL index, n	state_func	idx_down	idx_up	state_up	State_down	slope
1	1	0.04	0.0	0	0.0	0	25	0.0	0.0	0.0
2	26	23.45	0.0	1	0.0	25	34	0.0	0.0	0.0
3	35	56.13	0.0	2	0.5	34	38	0.0	0.5	1.53
4	39	77.24	1.0	3	0.5	38	43	0.5	0.5	0.0
5	44	110.24	0.0	4	0.0	43	48	0.5	0.0	-1.55
6	49	151.27	0.0	5	0.0	48	55	0.0	0.0	0.0
7	56	223.44	0.0	6	0.0	55	62	0.0	0.0	0.0
8	63	314.14	0.0	7	0.0	62	79	0.0	0.0	0.0

9	80	617.51	0.0	8	0.0	79	99	0.0	0.0	0.0
10	100	1100.00	0.0							

`slb2fin.pro` outputs the variable ‘func_fine’ with dimension [100]. This function corresponds to the fourth (from the top) O₃ trapezoid state function as depicted in Figure 56. When we look again at Table 6, we see that the trapezoid state function has a slope of 1.53 from pressure level index 35 to 39, then it has zero slope from pressure index 39 to 44 (this is the face of the trapezoid function) and a slope of -1.55 from pressure index 44 to 49.

The `calc_finv.pro` demonstrates how to implement Eq. 4 using IDL built-in routines.

3. IDL Routines

`PRO slb2fin`, numfunc, func_indx, func_apml, usehalftop, usehalfbot, presbot, air_pres, \$ func_fine

```

; -----
; PURPOSE: construct a trapezoid state function on 100 pressure levels
;           one at a time, using [ave_kern/*_func_indxs] hinge points
; -----
; INPUT:
; NAME      DESCRIPTION
; -----
; numfunc   number of trapezoid state functions above Earth surface at
;            retrieval scene
; func_indx  index values for trapezoid hinge points, starting at 1
; func_apml  amplitude of trapezoids, which is 1.0 for the
; usehalftop 0 is a trapezoid, 1 is a wedge
; usehalfbot 0 is a trapezoid, 1 is a wedge
; presbot    pressure [hPa] of bottom retrieval level = 1100.0
; air_pres    100-level retrieval pressure grid [air_pres]/100. in [hPa]

; OUTPUT:
; func_pfine  trapezoid state function on standard pressure level grid
;
;
; THIS FUNCTION IS CALLED BY CALC_FINV_MP.PRO
;
; -----
; Step 1: Construct the face of the trapezoid state function
; -----

state_face = fltarr(numfunc)

IF(usehalftop GT 0) THEN BEGIN
    state_face(0) = 0.5 * func_apml(0)
ENDIF ELSE BEGIN
    state_face(0) = func_apml(0)
ENDELSE

FOR n = 1, numfunc-2 DO BEGIN
    state_face(n) = 0.5 * ( func_apml(n) + func_apml(n-1) )

```

```

ENDFOR

IF (usehalfbot GT 0) THEN BEGIN
  state_face(numfunc-1) = 0.5 * func_ampl(numfunc-2)
ENDIF ELSE BEGIN
  state_face(numfunc-1) = func_ampl(numfunc-2)
ENDELSE

; -----
; Step 2: Calculate the state function on standard 100 level grid
; -----

FOR n = 0, numfunc-2 DO BEGIN
  idx_up = func_indx(n)-1
  idx_down = func_indx(n+1) - 1

  state_up = state_face(n)
  state_down = state_face(n+1)

  pres_up = alog(air_pres(idx_up))
  pres_down = alog(air_pres(idx_down))

  slope = (state_down - state_up)/(pres_down - pres_up)

  FOR L = idx_up, idx_down-1 DO BEGIN
    func_fine(L) = state_up + slope * (alog(air_pres(L)) - pres_up)
  ENDFOR
ENDFOR

func_fine(idx_down) = pres_down

END

```

```

PRO calc finv mp, num_func, func_indx, ret_nlev, htop, hbot, air_pres, $
  f_matrix, f_inv

; -----
; PURPOSE: calculates a scene-dependent transformation matrix (F_matrix) and
;           its inverse the using Moore-Penrose pseudoinverse technique
;           This matrix is scene-dependent because we use only those functions
;           and pressure levels that are above Earth surface at target scene.

; INPUT:
; Name      Description
; -----
; num_func  number of state functions above Earth surface
; func_indx trapezoid state function hinge-points
; ret_nlev  number of pressure levels (air_pres) above Earth surface
; htop      value in [ave_kern/*_func_hbot]
; hbot      value in [ave_kern/*_func_htop]
; air_pres  standard 100 level pressure grid [air_pres]/100 in hPa units

; OUTPUT:
; f_matrix  transformation matrix where each retrieval state function is
;           on the standard retrieval pressure grid
; f_inv     pseudoinverse matrix of f_matrix
; -----

```

```

; Step 1: calculate the transformation matrix: f_matrix
; -----
num_func = num_func_indx-1
f_matrix = FLTARR(num_func, ret_nlev)
FOR ifunc = 0, nfunc - 1 DO BEGIN
; Call slb2fin for one state function at a time setting the corresponding
; slbval = 1.0
slbval = FLTARR(num_func)
slbval(ifunc) = 1.0
fine = FLTARR(ret_nlev)
SLB2FIN, num_func_indx, func_indx, slbval, htop, hbot, $
1100., air_pres, fine
f_matrix(ifunc,*) = fine
ENDFOR
; -----
; Step 2: calculate the inverse of f_matrix using
; the Moore-Penrose pseudoinverse method
; -----
fftr = MATRIX_MULTIPLY(DOUBLE(f_matrix),DOUBLE(f_matrix),/BTRANS)
status=1L
finv1 = LA_INVERT(fftr,STATUS=status,/DOUBLE)
f_inv = MATRIX_MULTIPLY(finv1,DOUBLE(f_matrix))
END

```

References

- Clerbaux, C., Hadji-Lazaro, J., Hauglustaine, D., Mégie, G., Khattatov, B. and Lamarque, J.-F.: Assimilation of carbon monoxide measured from satellite in a three-dimensional chemistry-transport model, *Journal of Geophysical Research: Atmospheres*, 106(D14), 15385–15394, doi:10.1029/2000JD900682, 2001.
- Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados-Muñoz, M. J., Hannigan, J. W., Hase, F., Hassler, B., Huang, G., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S., Latter, B., Leblanc, T., Le Flochmoën, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M., Trickl, T., Weatherhead, E., Wespes, C., Worden, H. M., Vigouroux, C., Xu, X., Zeng, G., Ziemke, J., Helmig, D. and Lewis, A.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, *Elem Sci Anth*, 6, doi:10.1525/elementa.291, 2018.
- Iturbide-Sanchez, F., Liu, Q., Gambacorta, A., Barnet, C., Nalli, N. R., Tan, C. and Santos da Silva, S. R.: Using averaging kernels to study the vertical resolution of nucas temperature and water vapor, in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 33–35, IEEE, Fort Worth, TX., 2017.
- Levelt, P. F., Khattatov, B. V., Gille, J. C., Brasseur, G. P., Tie, X. X. and Waters, J. W.: Assimilation of MLS ozone measurements in the global three-dimensional chemistry

- transport model ROSE, *Geophysical Research Letters*, 25(24), 4493–4496, doi:10.1029/1998GL900152, 1998.
- Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T. and Chen, L.: Simultaneous assimilation of AIRS Xco₂ and meteorological observations in a carbon climate model with an ensemble Kalman filter: ASSIMILATION OF AIRS XCO₂, *Journal of Geophysical Research: Atmospheres*, 117(D5), n/a-n/a, doi:10.1029/2011JD016642, 2012.
- Maddy, E. S. and Barnet, C. D.: Vertical Resolution Estimates in Version 5 of AIRS Operational Retrievals, *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2375–2384, doi:10.1109/TGRS.2008.917498, 2008.
- Maddy, E. S., Barnet, C. D. and Gambacorta, A.: A Computationally Efficient Retrieval Algorithm for Hyperspectral Sounders Incorporating *A Priori* Information, *IEEE Geoscience and Remote Sensing Letters*, 6(4), 802–806, doi:10.1109/LGRS.2009.2025780, 2009.
- Nalli, N. R., Barnet, C. D., Reale, A., Tobin, D., Gambacorta, A., Maddy, E. S., Joseph, E., Sun, B., Borg, L., Mollner, A. K., Morris, V. R., Liu, X., Divakarla, M., Minnett, P. J., Knuteson, R. O., King, T. S. and Wolf, W. W.: Validation of satellite sounder environmental data records: Application to the Cross-track Infrared Microwave Sounder Suite: SATELLITE SOUNDER EDR VALIDATION, *Journal of Geophysical Research: Atmospheres*, 118(24), 13,628–13,643, doi:10.1002/2013JD020436, 2013.
- Pierce, R. B., Al-Saadi, J., Kittaka, C., Schaack, T., Lenzen, A., Bowman, K., Szykman, J., Soja, A., Ryerson, T., Thompson, A. M., Bhartia, P. and Morris, G. A.: Impacts of background ozone production on Houston and Dallas, Texas, air quality during the Second Texas Air Quality Study field mission, *Journal of Geophysical Research*, 114, doi:10.1029/2008JD011337, 2009.
- Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore; Hackensack, N.J., 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments: INTERCOMPARISON OF REMOTE SOUNDERS, *Journal of Geophysical Research: Atmospheres*, 108(D3), n/a-n/a, doi:10.1029/2002JD002299, 2003.
- Segers, A. J., Eskes, H. J., van der A, R. J., van Oss, R. F. and van Velthoven, P. F. J.: Assimilation of GOME ozone profiles and a global chemistry–transport model using a Kalman filter with anisotropic covariance, *Quarterly Journal of the Royal Meteorological Society*, 131(606), 477–502, doi:10.1256/qj.04.92, 2005.
- Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), *Remote Sensing*, 11(10), 1227, doi:10.3390/rs11101227, 2019.
- Smith, N. and Barnet, C. D.: CLIMCAPS Observing Capability for Temperature, Moisture and Trace Gases from AIRS/AMSU and CrIS/ATMS, preprint, *Gases/Remote Sensing/Data Processing and Information Retrieval.*, 2020.
- Strow, L. L., Hannon, S. E., De Souza-Machado, S., Motteler, H. E. and Tobin, D.: An overview of the AIRS radiative transfer model, *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 303–313, doi:10.1109/TGRS.2002.808244, 2003.
- Susskind, J., Barnet, C. D. and Blaisdell, J. M.: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE TGRS*, 41, 390–409, 2003.

- Susskind, J., Blaisdell, J. M. and Iredell, L.: Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm, *Journal of Applied Remote Sensing*, 8(1), 084994, doi:10.1117/1.JRS.8.084994, 2014.
- Yudin, V. A.: Assimilation of the 2000–2001 CO MOPITT retrievals with optimized surface emissions, *Geophysical Research Letters*, 31(20), doi:10.1029/2004GL021037, 2004.

Chapter 6: Quality Control

CLIMCAPS uses Quality Control (QC) flags to intuitively communicate which vertical and spatial observations of temperature, water vapor (H₂O), and trace gases are appropriate to use. We will describe how to access, interpret, and use the QC flags inside CLIMCAPS netCDF file.

Section 1: Vertical Quality Control Flags

In the CLIMCAPS netCDF file, users can access the QC flags in the fixed pressure grid for each of the retrieved and derived variables by accessing fields that end in ***_qc**. CLIMCAPS defines retrievals as successful if all steps of the retrieval successfully executed; if any steps do not pass or if they fail any additional internal quality checks, then the retrieval is flagged as having failed. All variables inside the netCDF file that end in ***_qc** have the exact same values because CLIMCAPS V2 employs a single QC schema for all retrieval variables. CLIMCAPS being a step-wise retrieval procedure, we argue that if T(p) fails, then all subsequent retrieval variables should also be flagged as failed. For example, **co_mol_lay_qc** has identical values to those in the **o3_mol_lay_qc** variable. In the future, we may consider adopting variable-specific QC flags, so we recommend using the variable-specific QC flag in your code. For example, if you are working with **co_mol_lay**, then use the **co_mol_lay_qc** variable. The QC flag has three possible values, which are 0, 1, and 2. To improve understanding of QC, we encourage users to interpret these respective values as best, good, and rejected (Table 4). Note that in addition to the ***_qc** variables, there are error (***_err**), degrees of freedom (***_dof**), and averaging kernel matrix variables inside the netCDF file, which are also important indicators of quality and information content.

Table 7: Description of quality control values and their appropriate use.

Value	Meaning	Appropriate Use
0	Best	Best QC retrievals can be used without reservation following the guidance of the variable-specific application guides. We recommend that users develop a general understanding of the available information content of the variable to ensure they are appropriately interpreting atmospheric state variables. Best QC retrievals can be used for <i>in situ</i> measurement comparison, data assimilation, or another research application.
1	Good	When combined with the best QC retrievals, good QC retrievals increase the yield of available profiles and thus are useful for applications that require a large sample of retrievals. Good QC retrievals are appropriate for applications where measurements are spatially or temporally aggregated. It is recommended that users analyze good retrievals alongside measures of variable information content, which includes degrees of freedom (DOF; *_dof) and the Averaging Kernel Matrix (AKM; ave_kern/*_ave_kern), and

		errors (*_err). These are described in detail in variable-specific application guides.
2	Rejected	We do not recommend using rejected retrievals unless the region is data sparse. These are cases where it may be more appropriate to use the CLIMCAPS MW-only retrievals (mw/mw_air_temp , mw/mw_h2o_vap_mol_lay). If using IR+MW retrievals, it is strongly recommended that rejected retrievals are analyzed alongside measures of information content, which include DOFs, AKMs, and errors. Additionally, rejected retrievals should also be evaluated in context of cloud clearing parameters (aux/etarej , aux/ampl_eta , and aux/aeff_end), which are described later in this user guide.

Figure 58 illustrates the global and zonal yield of QC by height. Measurements that are below the surface pressure are not included in Figure 58 or the discussion below. However, note that measurements below the surface will automatically have a QC flag value of 2 (rejected) in the netCDF file. We calculated our yield fraction by dividing the number of observations with a specific quality flag with the total number of observations at the given pressure level.

Clouds are one of the primary reasons why retrievals fail. For this reason, retrievals in the stratosphere have a much higher yield than in the troposphere, which is reflected in the sharp discontinuity above the tropopause of each zone.

Across all zones, the yield is nearly 1.0 above 200 hPa for ‘best’ and the combined ‘best’ and ‘good’ (best+good) quality retrievals. There is greater variability between 1000 hPa and 800 hPa across each of the latitude zones shown in Figure 58. For example, the lowest yield occurs in the southern polar region (60°S-90°S), which between 600-200 hPa has a rejection rate of 0.30. Combining ‘best’ and ‘good’ retrievals increases the yield significantly in the tropics (30°N-30°S).

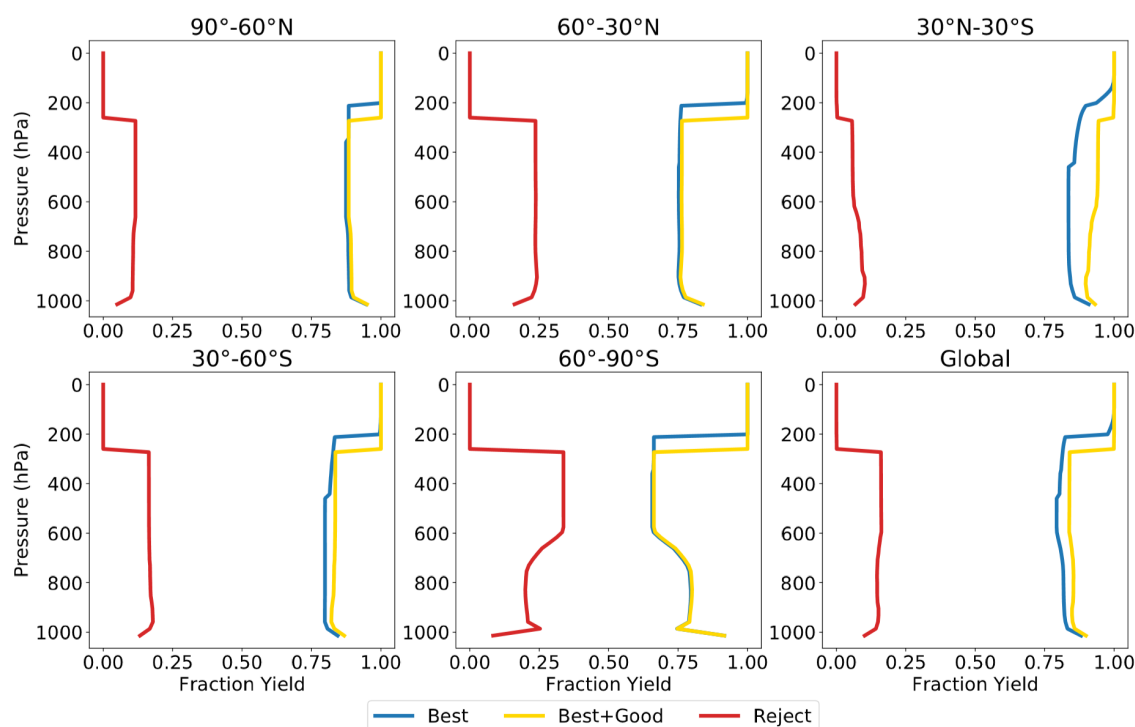


Figure 59: Fraction of retrievals with a given QC flag for CLIMCAPS-SNPP T(p) retrievals (**air_temp_qc**) from full spectral resolution cloud cleared CrIS radiances. The results are shown for five latitudinal zones and for the full global retrieval. 'Best' quality data has a QC flag value of 0, 'best+good' has a value of 0 or 1, and rejected data has a value of 2. Note that while each variable has a corresponding *_qc field, these values do not vary between CLIMCAPS state and derived variables. The example presented here is from CLIMCAPS-SNPP retrievals on 1 April 2016.

Section 2: Footprint Quality Control

Users may also wish to use a single QC metric for the entire footprint, rather than by pressure level. For this application, we recommend the QC approach adopted by the NOAA-Unique Combined Atmospheric Processing System (NUCAPS) for the National Weather Service (NWS) that we summarized in a quick guide [<https://weather.msfc.nasa.gov/nucaps/qg/NUCAPS-QF-quick-guide.pdf>]. NUCAPS is a sister algorithm to CLIMCAPS, specifically designed for real-time monitoring of hazardous weather.

This NUCAPS QC method was developed to facilitate forecaster interpretation during severe weather events and uses a visually intuitive, “stoplight” color coding approach. In this method, footprints are labeled green when both microwave and infrared retrievals pass (clear sky/partly

cloudy conditions), yellow when the MW-only passes (cloudy conditions), and red when both the combined IR+MW and the MW-only steps fail (precipitating conditions).

An example of this QC method is shown in Figure 60, which shows a screen capture of NUCAPS NOAA-20 from the NWS AWIPS-II visualization software.

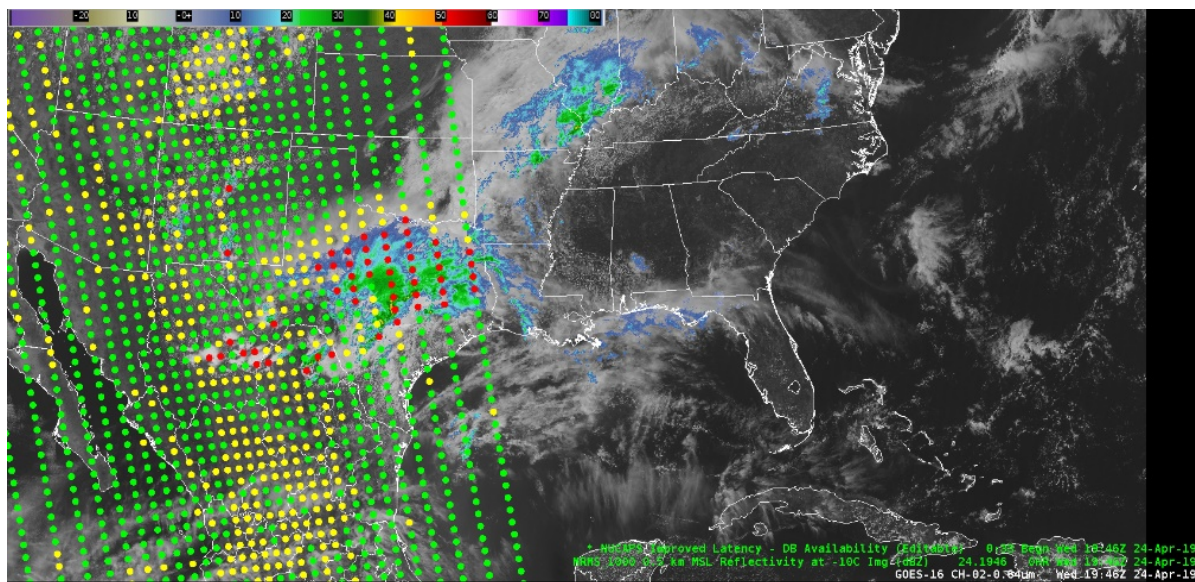


Figure 60: Example of a footprint-level QC approach used in the NUCAPS algorithm by the National Weather Service (NWS). Green retrievals indicate the IR+MW step successfully passed, yellow retrievals where the MW-only step passed while the IR+MW step failed, and red indicates both steps were rejected. The example is taken from a screen capture from AWIPS-II of NUCAPS from a NOAA-20 overpass of convection at 19:45 UTC on April 24, 2019.

The NUCAPS single footprint QC method can also be implemented in CLIMCAPS by reading the **aux/ispare_2** in the netCDF file. Figure 61 shows the zonal yield fraction for each of these QC flags.

Using the NUCAPS QC method, the highest yield (0.87) of retrievals where MW+IR passed is in the northern polar region (90°N-60°N) and lowest yield (0.66) is in the southern polar region (60°S-90°S). Globally, the percent yield is roughly 0.80, with only a 0.03 rate of rejection.

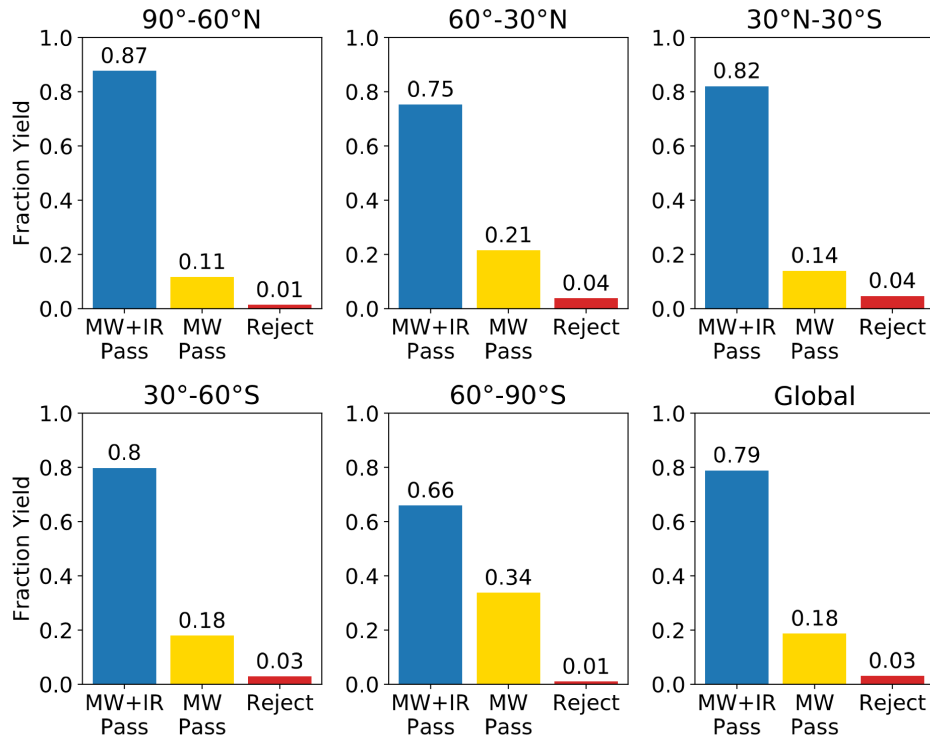


Figure 61: Fraction of retrievals with a given QC flag for CLIMCAPS-SNPP T(p) retrievals (**aux/ispare_2**) from full spectral resolution cloud cleared CrIS radiances. The results are shown for five latitudinal zones and for the full global retrieval. 'MW+IR Pass' data has a bit flag of zero, 'MW-only Pass' has a value of 1, and 'Reject' has a value of 9. Results presented here are for CLIMCAPS-SNPP retrievals on 1 April 2016.

The values inside the **aux/ispare_2** field have three values: 0, 1, and 9, which correspond to MW+IR Pass, MW Pass, and Reject. These numbers are not sequential because they are encoded as 8-bit binary values to diagnose which components of the retrieval failed or succeeded. However, the values can be read as integers and do not need to be read bitwise in order to interpret their meaning.

1. Data filtering options

In addition to quality flags, users may wish to filter CLIMCAPS footprints by time of day or scene type, such as during sunlight, over land or ocean, and clear or cloudy scenes. Furthermore, users may wish to filter the vertical measurements into the troposphere or stratosphere. Below, we include some data filtering options. Note that for all options below, *i* and *j* respectively refer to the footprint indices along the *atrack* and *xtrack*.

– *Ascending versus Descending*

- **asc_flag (atrack):** The ascending flag is useful for determining whether a granule is from an ascending or descending orbit. If $\text{asc_flag}(i) = 1$ then the retrieval is from the ascending orbit, which means it has a 01:30 pm local overpass time. If $\text{asc_flag}(i) = 0$

then the retrieval is from the descending orbit and thus at the 01:30 am local overpass time.

- **lat(atrack,xtrack):** Changes in latitude between scanlines is also a helpful indicator of the orbit direction. If $\text{lat}(i+1,j) > \text{lat}(i,j)$ then the granule is in ascending (01:30 pm) orbit and if $\text{lat}(i+1,j) < \text{lat}(i,j)$ then granule is in the descending (01:30 am) orbit.

– *Sun versus No Sun*

- **sol_zen (atrack,xtrack):** The solar zenith angle is a function of latitude, the local time, and day of the year. The solar zenith angle is useful for determining if the sun is above or below the horizon at a specific location, thus it may be used to classify a footprint as being exposed to sunlight or not. For $\text{sol_zen}(i,j) = 0$, the sun is directly above the footprint, and for $\text{sol_zen}(i,j) = 90$ the sun is at the horizon at the time of satellite measurement. If $\text{sol_zen}(i,j) > 90$, then sun is below the horizon and thus have no impact on the radiance measurement or retrieval. In CLIMCAPS, we use a threshold of $\text{sol_zen}(i,j) = 89.9^\circ$ to determine if solar reflectivity should be accounted for or not.

– *Land versus Ocean*

- **land_frac (atrack,xtrack):** The land fraction measures how much a footprint falls over land as opposed to water. For example, if $\text{land_frac}(i,j) \geq 0.75$ (thus 75%) then the retrieval footprint is mostly over land. One can consider a range of $0.75 > \text{land_frac}(i,j) \geq 0.25$ (thus 25-75%) for a retrieval footprint to be over coastlines. If $\text{land_frac}(i,j) < 0.25$ (thus 25%) then the retrieval footprint is mostly over ocean. In CLIMCAPS V2 we do not distinguish between land or ocean during retrieval. Instead, for a footprint with mixed surface types, CLIMCAPS simply makes a weighted average (as determined by the fraction of each surface type) of the land and ocean emissivity spectra. In applications, you may wish to clearly distinguish land from ocean, so you may define your own land_frac thresholds.

– *Clear-sky versus Cloudy*

- **aux/cldfrac_tot(atrack,xtrack):** cldfrac_tot estimates the total cloud fraction over the retrieval footprint (3 x 3 fields of view; ~50 km at nadir, ~150 km at edge of scan). CLIMCAPS employs cloud clearing, which allows successful retrievals in up to 90% cloudy conditions. Your application may need to distinguish between clear and cloudy scenes, so you can use $\text{cldfrac_tot}(i,j) > 0.10$ (thus more than 10% cloud cover) as ‘cloudy’ and $\text{cldfrac_tot}(i,j) \leq 0.10$ as ‘clear’. You can, of course, vary these values according to the requirements of your application.
- **aux/cldfrac_500(atrack,xtrack):** cldfrac_500 is the total cloud fraction below 500 hPa over the retrieval footprint. This metric is useful for identifying scenes with lower tropospheric clouds.
- **cld_frac(atrack,xtrack, fov, cld_lay):** Like aux/cldfrac_tot , cld_frac is the total cloud fraction but instead it is over the instrument field of view (using the fov index) and not the CLIMCAPS footprint (3 x 3 fovs). **cld_frac** is available as two cloud layers using the cld_lay index, where $\text{cld_lay} = 1$ is the lower cloud layer and $\text{cld_lay} = 2$ is the upper

cloud layer over the instrument field of view. The interpretation of the values is the same as that of *aux/cldfrac_tot*.

- **aux/etarej(atrack,xtrack):** Rather than inspect the cloud fraction, you may wish to filter retrievals based on cloud uncertainty. This metric is the cloud clearing radiance error in brightness temperature units [Kelvin]. CLIMCAPS calculates it as radiance residual, or the difference between the simulated clear-sky radiance estimate and the retrieved cloud cleared radiance at a retrieval scene. **etarej**, thus, quantifies the quality of cloud clearing by indicating how well the cloud-cleared radiance represents the clear-sky state around clouds. **etarej** is one of the QC criteria employed in determining whether a retrieval failed or not (using a threshold of 1.5K). Smaller values of **etarej** indicate successful cloud clearing and a high confidence in the removal of the radiance signal from clouds. Higher values of **etarej** indicate that the cloud cleared radiance channels are ‘contaminated’ by cloud radiative effects that we were unable to remove during the cloud clearing step. We recommend that you use **etarej** to identify CLIMCAPS retrievals with cloud contamination, or uncertainty due to clouds.
- **aux/ampl_eta(atrack,xtrack):** The amplification factor (**ampl_eta**) quantifies how much the random instrument noise (NEN or NEdT in units Kelvin) was amplified ($\text{ampl_eta}(i,j) > 1$) or damped ($\text{ampl_eta}(i,j) < 1$) as a result cloud clearing.
- **aux/aeff_end(atrack,xtrack):** The effective amplification factor (**aeff_end**) is a compound metric that combines **random** (instrument noise scaled by **ampl_eta**) and **systematic** (spectral correlation) cloud clearing uncertainty.

– *Troposphere versus Stratosphere*

- **tpause_pres(atrack,xtrack):** The tropopause pressure is useful for stratifying the troposphere from the stratosphere.

Chapter 7: CLIMCAPS Infrared Channel Selection

CLIMCAPS V2 is configured to generate sounding products (Level 2 files) from instruments on three different satellite platforms. These are: AIRS/AMSU on Aqua, CrIS/ATMS on Suomi-NPP and CrIS/ATMS on JPSS-1. CLIMCAPS retrieves multiple atmospheric variables from a single cloud-cleared infrared (IR) radiance measurement using a sequential Optimal Estimation approach (Smith and Barnett, 2019, 2020).

CLIMCAPS uses subsets of channels to retrieve atmospheric variables one at a time in a sequence as depicted in the flow-diagram. These subsets of channels, each an order of magnitude lower than the total number of channels, significantly reduces retrieval execution time. CLIMCAPS retrieves the full set of atmospheric state variables within ~150 ms per cloud-cleared scene, or 6 min for a 6 min Level 1b granule. The number of channels we select for each variable depends on the measurement signal-to-noise and degree of interference from other variables. Below we depict the IR channel subsets for each retrieval variable and each instrument configuration – AIRS, CrIS nominal spectral resolution (CrISNSR) and CrIS full spectral resolution (CrISFSR). The CLIMCAPS IR channel selection methodology is detailed elsewhere (Gambacorta and Barnett, 2011, 2013) and we briefly outline it here.

The channel sets we present here is for CLIMCAPS V2 and can easily be changed or updated in future versions as we improve the system across instruments. The channel lists are defined in FORTRAN namelists, not in the source code itself.

1. CLIMCAPS channel subsets for AIRS and CrIS

In Table 8 we summarize the main instrument differences.

AIRS (Atmospheric Infrared Sounder; Aumann et al., 2003) is a grating interferometer with 2378 spectral channels, each a pair of detectors – “A” and “B”. AIRS has 17 arrays of detectors and it is possible for individual detectors to be bad or compromised. If both detectors are good then their spectral values are averaged, and their noise reduced to the square root of the two channel noise values. AIRS detector characteristics can change over time (colloquially known as “popping”) and especially where events caused the instrument to shut-down and warm up again. AIRS spectral resolution changes with wavenumber but is roughly equal to the wavenumber, ν , divided by 1200, and the spectral equivalent to $\nu/2400$. Each AIRS channel and each field-of-view (FOV) can be considered an independent measurement. There are 9 x FOVs collocated within each AMSU footprint. The CLIMCAPS cloud clearing step combines these 9 FOVs into a single cloud-cleared radiance, known as a field-of-regard (FOR) with coarser spatial resolution, reduced instrument noise and the radiative effects of clouds removed from each cloud-sensitive channel.

CrIS (Cross-track Infrared Sounder; Glumb et al., 2002; Strow et al., 2013) is an interferometer that measures 9 x FOVs measured simultaneously for three spectral bands – short-wave (645–1210 cm^{-1}), mid-wave (1210–2000 cm^{-1}) and long-wave (2000–2760 cm^{-1}). CrIS radiances are unapodized in the Level 1 product files, hence the presence of duplicate ‘guard’ channel at the ends of each band. In CLIMCAPS, we apply Hamming apodization according to the method

described in (Barnet et al., 2000; Gambacorta and Barnet, 2011). In Tables 8 and 9 below, we exclude guard channels from the total number of channels reported for CrIS.

Table 8: Overview of the two hyperspectral infrared (IR) instruments on low-Earth orbiting satellite platforms that CLIMCAPS is configured for. CrIS on SNPP was at first transmitted at nominal spectral resolution (NSR) but later switched to full spectral resolution (FSR) so we list statistics for both. These instruments measure the top of atmosphere radiance with spatial footprints known as (FOV) and CLIMCAPS retrieves soundings from every 3 x 3 array of FOVs known as the Field of Regard (FOR). Instrument noise is given in brightness temperature units as the noise equivalent delta temperature (NEDT) for a scene at 250K. This table was originally published in (Smith and Barnet, 2019). Note that we exclude mention of the two CrIS duplicate ‘guard’ channels for each spectral band.

Instrument Type	AIRS ¹ Grating	CrIS ³ Interferometer
Satellite	Aqua	SNPP, NOAA-20
Launch	2002/05/04	2011/10/28, 2017/11/18
Local Overpass Time	13:30	13:30
Altitude (km)	705	824
Mass (kg)	177	147
Period (min)	98.8841	101.4978
Orbits/day	14.5625	14.1875
FOV (deg)	1.1100	0.963
FOV (km)	13.5	14
# FOV per FOR 50km@nadir	9	9
# FOR ⁴ per day	30 x 10800 = 324,000	30 x 10800 = 324,000
Total # IR spectral channels		
LW ⁵ band (645–1210 cm ⁻¹)	2378	NSR: 1305, FSR: 2211
MW ⁶ band (1210–2000 cm ⁻¹)	1262	NSR: 713, FSR: 713
¹⁾	602	NSR: 433, FSR: 863
SW ⁷ band (2000–2760 cm ⁻¹)	514	NSR: 159, FSR: 865
¹⁾		
Spectral sampling (cm ⁻¹)	v/2400, 0.25 to 1.07	NSR: 0.625, 1.25, 2.5 FSR: 0.625 (all bands)
Apodization Type	n/a	Hamming (0.9/OPD)
Apodized Resolution (cm ⁻¹)	v/1200, 0.5 to 2.3	NSR: 1.125, 2.25, 4.5 FSR: 0.75 (all bands)
Noise characteristic		
NEDT (T=250K) @ 700 cm ⁻¹		
NEDT (T=250K) @ 1400 cm ⁻¹	0.23	NSR, FSR: 0.05
NEDT (T=250K) @ 2400 cm ⁻¹	0.08	NSR: 0.05, FSR: 0.07
	0.14	NSR: 0.2, FSR: 0.5

¹Atmospheric InfraRed Sounder; ²Infrared Atmospheric Sounding Interferometer; ³Cross-track Infrared Sounder; ⁴Field of Regard is the technical term but we refer to a retrieval footprint as the “datum” in this paper ⁵Longwave infrared; ⁶Midwave infrared; ⁷Shortwave infrared

Both AIRS and CrIS have 30 FORs along each scanline, 10,800 scanlines per day, and thus a total of 324,000 FORs from which CLIMCAPS retrieves atmospheric state variables. In Table 9 we summarize the number of spectral channels selected for each retrieval variable according to those available in AIRS, CrISNSR and CrISFSR. Even though CrISFSR instrument noise varies slightly between SNPP and JPSS1, we use the exact same channel sets for both. This is something we can revise in future, to test if JPSS1/CrISFSR could benefit from its own optimized channel sets.

Table 9: Spectral channel summary for Aqua/AIRS, Suomi-NPP CrIS normal spectral resolution (CrISNSR), Suomi-NPP CrIS full spectral resolution (CrISFSR), and JPSS-1 (or NOAA-20) CrISFSR, and eight CLIMCAPS retrieval variables. These channel totals are for apodized CrIS radiances, excluding count of the ‘guard’ channels. For each set of channels, we list the average degrees of freedom (DOF) achieved within CLIMCAPS for a global day of FORs. This table is an extract from one published in (Smith and Barnett, 2020).

	Aqua/AIRS Total=2387		SNPP/CrIS NSR Total=1305		SNPP/CrIS FSR Total=2211		JPSS1/CrIS FSR Total=2211	
	nch	DOF	nch	DOF	nch	DOF	nch	DOF
Temperature	134	6.3	86	3.5	120	3.0	120	3.0
Water Vapor (H ₂ O)	46	2.7	62	2.2	66	1.7	66	1.7
Ozone (O ₃)	40	2.0	53	2.3	77	1.9	77	1.9
Carbon Monoxide (CO)	36	0.7	27	0.2	35	0.8	35	0.8
Methane (CH ₄)	65	1.0	55	0.6	84	0.7	84	0.7
Carbon Dioxide (CO ₂)	61	0.7	53	0.9	54	0.8	54	0.8
Nitric Acid (HNO ₃)	14	0.3	28	0.3	30	0.1	30	0.1
Nitrous Oxide (N ₂ O)	58	1.2	24	0.8	21	0.3	21	0.3
Sulfur Dioxide (SO ₂)	60	0.02	24	1e-3	31	6e-4	31	7e-4

Figures 62 through 65 below depict the channel wavenumbers (cm⁻¹) for each instrument and each CLIMCAPS retrieval variable. Note that some retrieval variables share spectral channels. This is especially pronounced for cloud clearing (CC) and CTP/F (cloud top pressure and fraction). While this spectral overlap is deliberate in these two variables, it is may not be ideal in other cases. Figures 62-64 show the spectral channels that are used in the retrieval of a single variable, and in Figures 66 and 67 we show those that are shared between two or more retrieval variables. In Figures 68 and 69 we expand on these shared channels by depicting which retrieval variables share which channels.

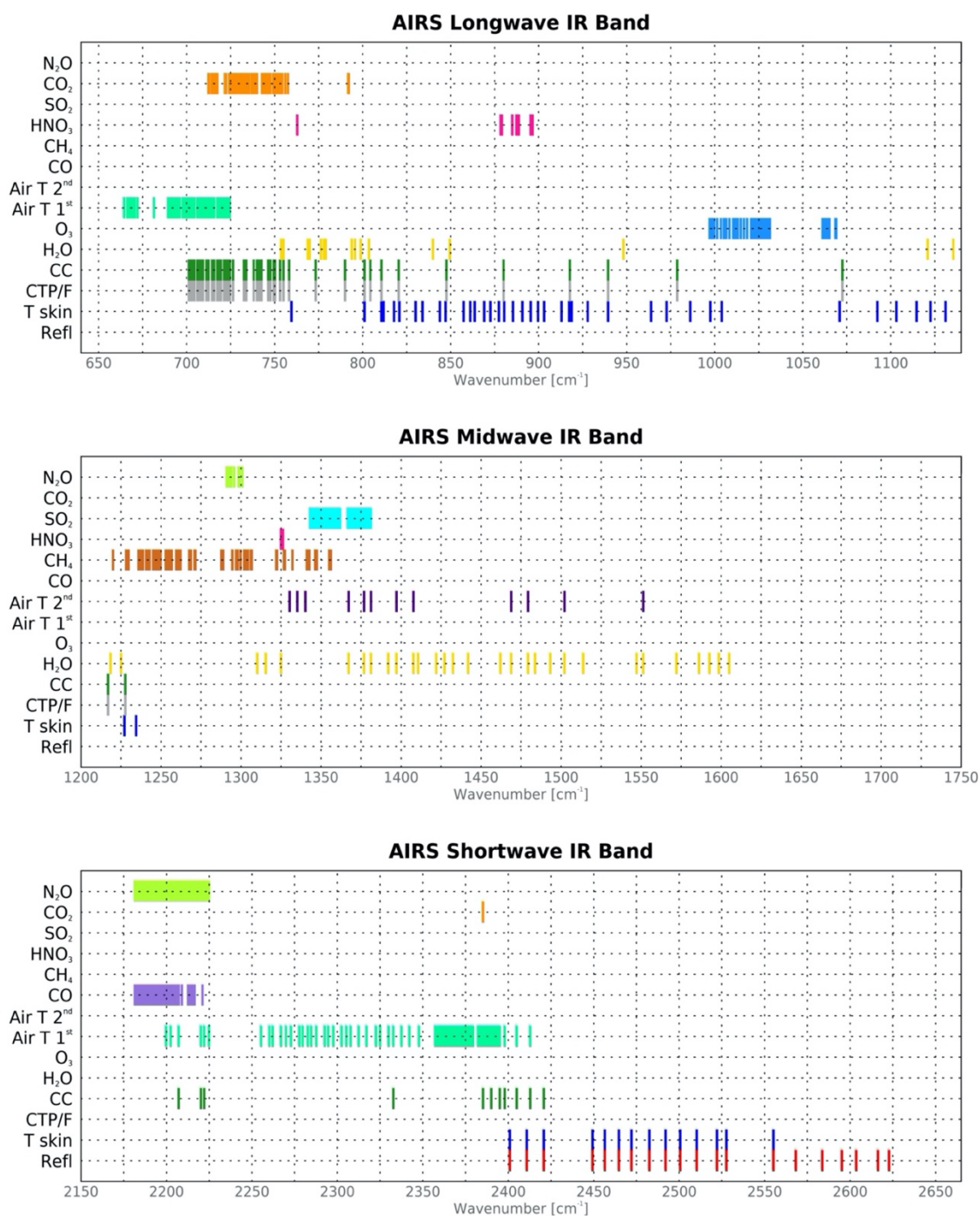


Figure 62: CLIMCAPS V2 channel sets for AIRS on Aqua in the (top) long-wave infrared (IR) band, (middle) mid-wave IR band and (bottom) short-wave IR band. Each item on the y-axis represents a CLIMCAPS retrieval variable with vertical lines indicating the channel wavenumbers for each variable. From top to bottom there is, Nitrous Oxide (N₂O), Carbon Dioxide (CO₂), Sulphur Dioxide (SO₂), Nitric Acid (HNO₃), Methane (CH₄), Carbon Monoxide (CO), atmospheric temperature (Air T) first and second passes (see flow diagram), Ozone (O₃), Water Vapor (H₂O), cloud clearing (CC), cloud top pressure (CTP) and cloud top fraction (CTF), Earth surface skin temperature (T skin) and Reflectivity (Refl).

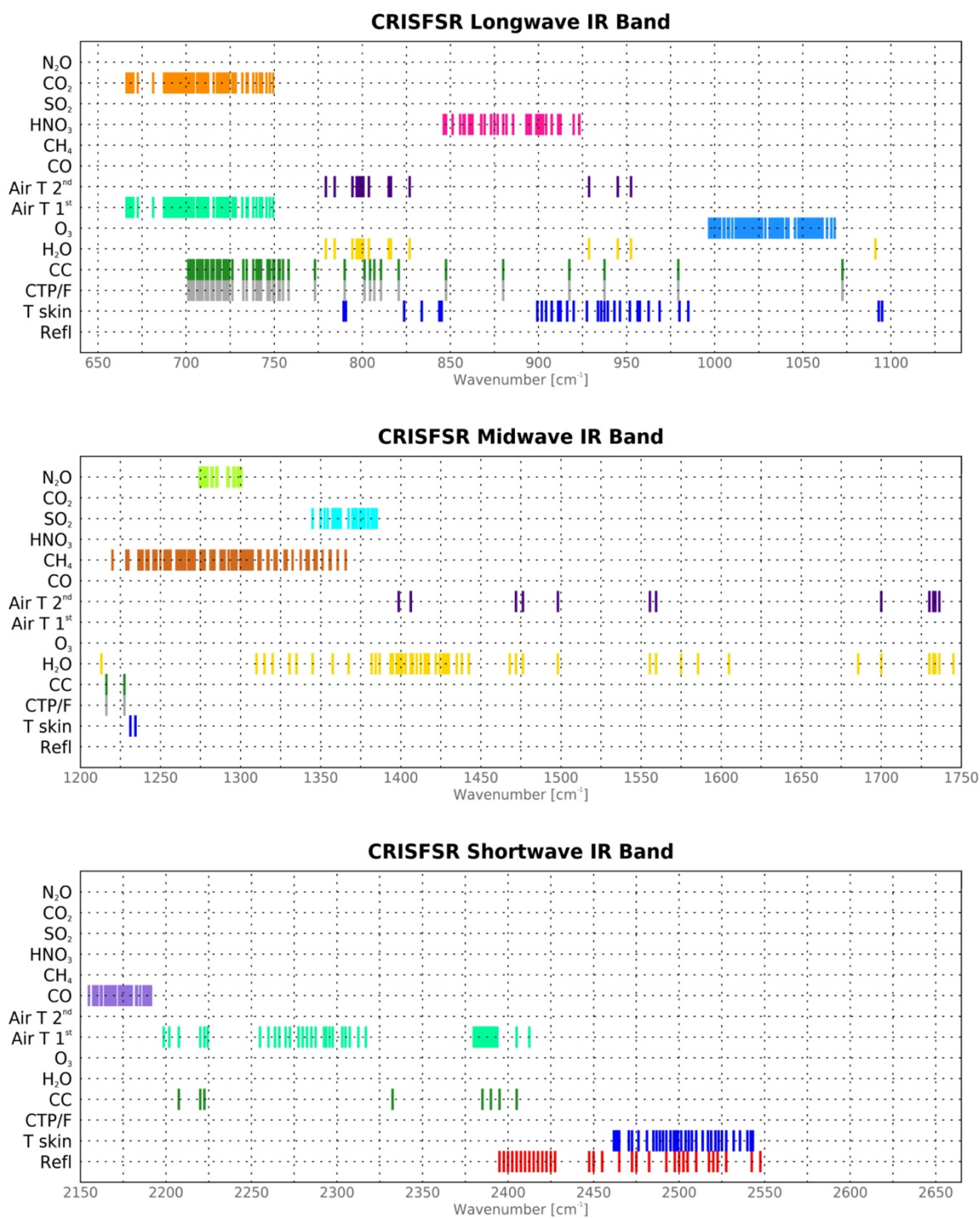


Figure 63: Same as Figure 62 but for CrIS full spectral resolution (FSR) on SNPP and JPSS1 (NOAA20).

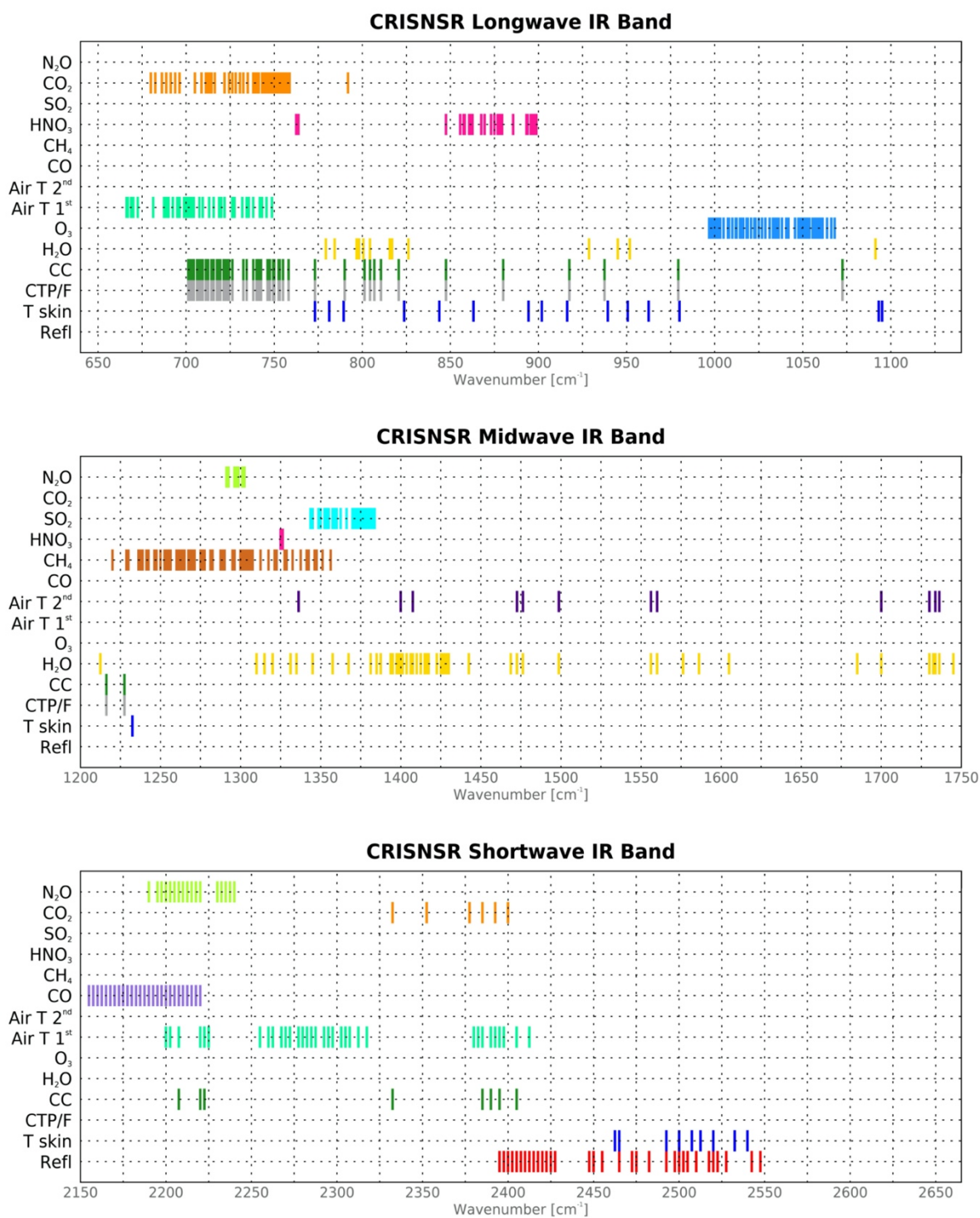


Figure 64: Same as Figure 62 but for CrIS normal spectral resolution (NSR) on SNPP.

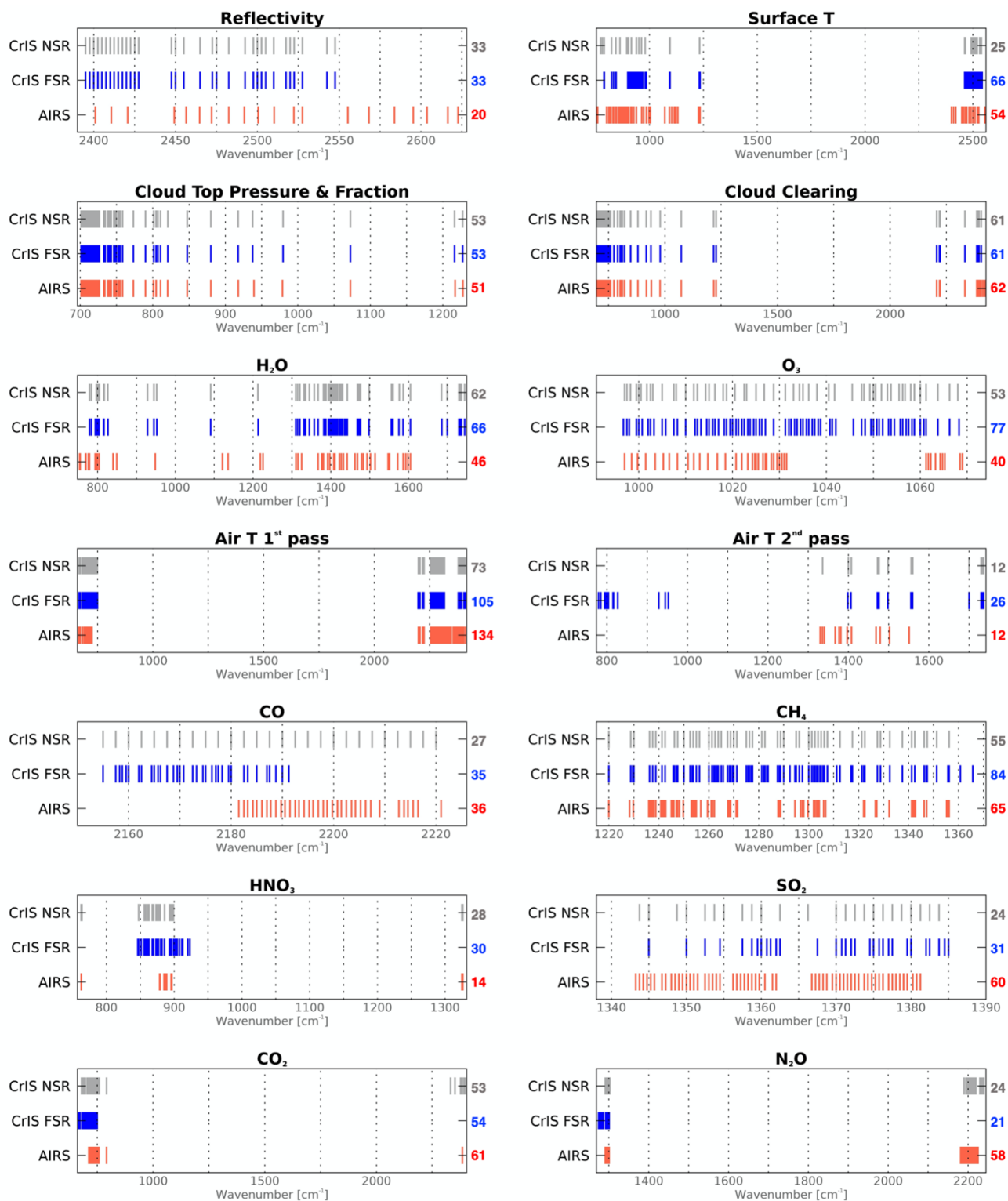


Figure 65: Same as Figures 62 through 64, but instead summarized according to retrieval variables with the y-axis items the three instrument configurations. On the right-hand y-axis of each panel, we list the total number of channels for selected for each instrument and target variable.

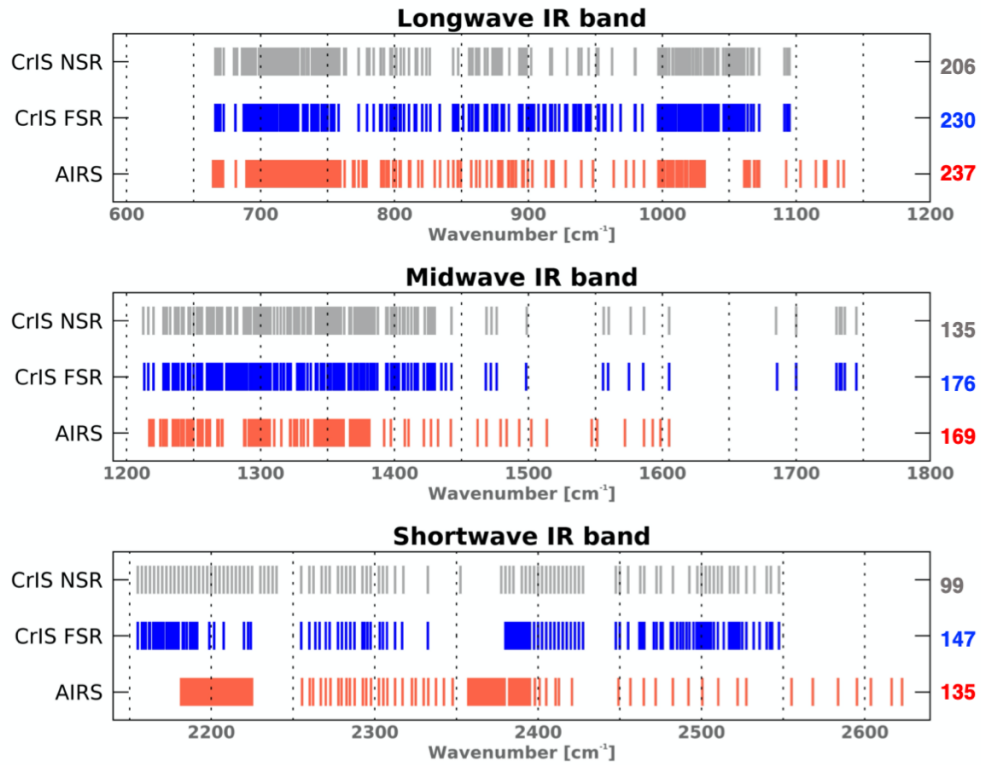


Figure 66: Wavenumbers of channels used in the retrieval of only one atmospheric variable, i.e., the set of unique spectral channels for the three infrared bands, (top) long-wave, (middle) mid-wave and (bottom) short-wave. We depict these for (grey) CrIS NSR, which has 440 unique channels, (blue) CrIS FSR with 553 unique channels and (red) AIRS with 541 unique channels. On the right-hand side of each panel, we list the total number of unique channels for each instrument and each spectral band.

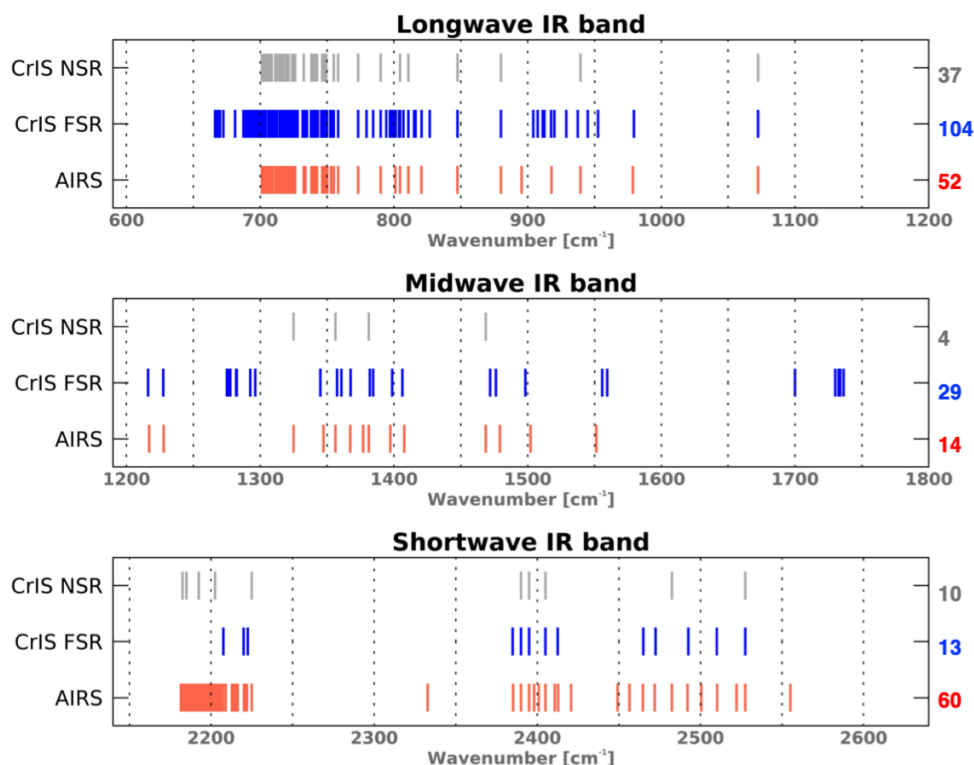


Figure 67: Same as Figure 66 but depicting the channels shared by two or more retrieval variables for each instrument and IR spectral band. CrIS NSR has a total of 52 channels used in multiple CLIMCAPS retrievals, CrIS 146, and AIRS 126.

We depict the CLIMCAPS retrieval variables that share spectral channels (Figures 66 and 67) and we do so for AIRS (Figure 68) and CrIS FSR (Figure 69). With this we wish to highlight some of the differences in our CLIMCAPS implementation for AIRS and CrIS. We note that spectral channels are sometimes shared between temperature (T_1 or T_2), skin temperature (T_{skin}), Nitric Acid (HNO_3), Carbon Dioxide (CO_2), Water Vapor (H_2O), Cloud Top Pressure (CTP) and Cloud Clearing (CCR). The number and wavenumbers of the shared channels vary between instruments and we recognize that this may affect continuity. We will revisit our CLIMCAPS channel sets, optimize where possible and work any updates into V3 (due for release in 2021).

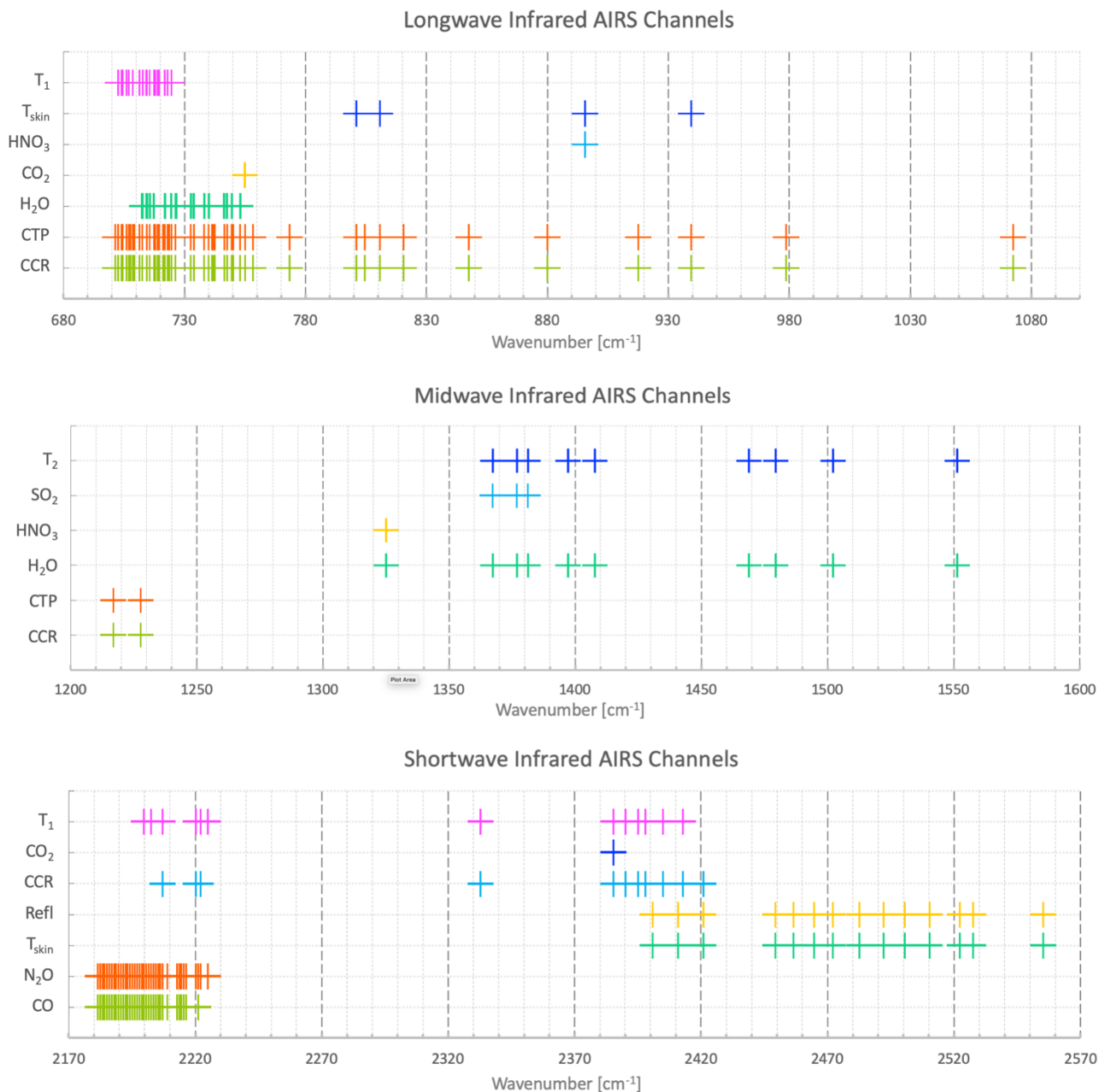


Figure 68: A break-down of the CLIMCAPS spectral channels used in two or more retrievals from AIRS measurements. The number and target of the shared channels vary in the (top) long-wave, (middle) mid-wave and (bottom) short-wave spectral bands.

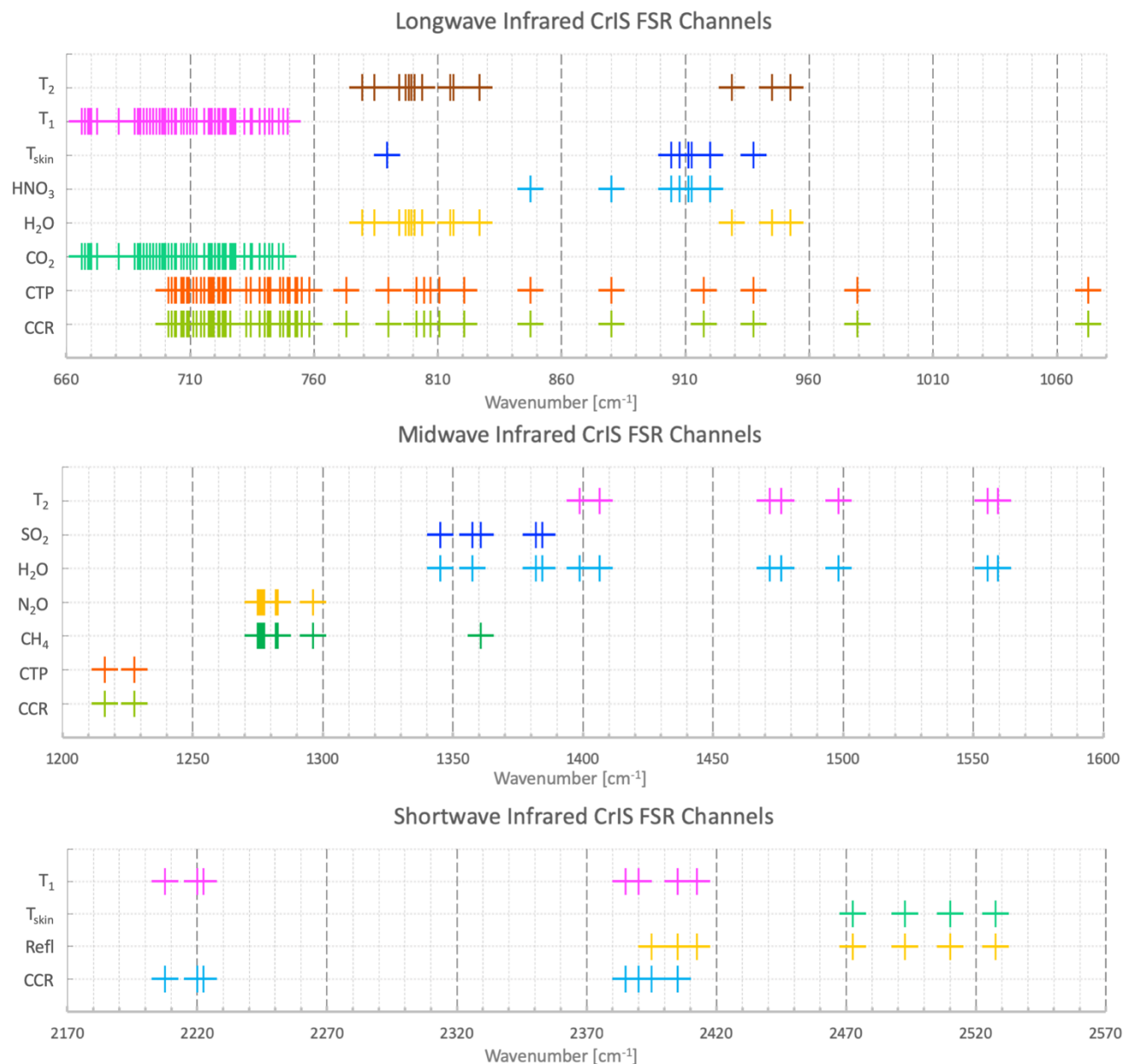


Figure 69: Same as Figure 64 but for CrIS full spectral resolution (FSR).

2. CLIMCAPS channel selection method

In CLIMCAPS we do not use all available IR channels to retrieve atmospheric variables. Instead, we select a subset of channels from the hundreds available according to instrument signal-to-noise estimates for each target variable. We reduce the number of spectral channels for at least two reasons; (i) they greatly speed up retrieval by reducing the dimension of the weighting function matrices (Jacobians) used in the optimal-estimation retrieval step (Smith and Barnett, 2020), (ii) they stabilize the signal-to-noise ratio of the measurements for a target variable to allow retrieval in a wide range of atmospheric conditions across the globe in all seasons.

We use the channel selection method first developed by the AIRS Science Team (Susskind et al., 2003) and later adopted by NOAA. This method differs from a purely mathematical approach in that it considers, not only the signal-to-noise with respect to the target variable, but also the signal-to-noise from background variables. The channel set we select maintains a robust information content for the target variable in all conditions, while minimizing the interference from background state variables. Gambacorta and Barnett (2011, 2013) described the method in detail.

References

- Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L. and Susskind, J.: AIRS/AMSU/HSB on the aqua mission: design, science objectives, data products, and processing systems, *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 253–264, doi:10.1109/TGRS.2002.808356, 2003.
- Barnett, C. D., Blaisdell, J. M. and Susskind, J.: An analytical transformation for use in computation of interferometric spectra for remote sensing applications, *IEEE TGRS*, 38, 169–183, 2000.
- Gambacorta, A. and Barnett, C. D.: Methodology and information content of the NOAA NESDIS operational channel selection for the Cross-Track Infrared Sounder (CrIS), edited by N. E. S. United States Data, and Information Service, [online] Available from: <https://repository.library.noaa.gov/view/noaa/1272>, 2011.
- Gambacorta, A. and Barnett, C. D.: Methodology and Information Content of the NOAA NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS), *IEEE Transactions on Geoscience and Remote Sensing*, 51(6), 3207–3216, doi:10.1109/TGRS.2012.2220369, 2013.
- Glumb, R. J., Jordan, D. C. and Mantica, P.: Development of the Crosstrack Infrared Sounder (CrIS) sensor design, edited by M. Strojnik and B. F. Andresen, pp. 411–424, San Diego, CA., 2002.
- Smith, N. and Barnett, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), *Remote Sensing*, 11(10), 1227, doi:10.3390/rs11101227, 2019.
- Smith, N. and Barnett, C. D.: CLIMCAPS Observing Capability for Temperature, Moisture and Trace Gases from AIRS/AMSU and CrIS/ATMS, preprint, *Gases/Remote Sensing/Data Processing and Information Retrieval*, 2020.
- Strow, L. L., Motteler, H., Tobin, D., Revercomb, H., Hannon, S., Buijs, H., Predina, J., Suwinski, L. and Glumb, R.: Spectral calibration and validation of the Cross-track Infrared Sounder on the Suomi NPP satellite: CrIS Spectral Calibration, *Journal of Geophysical Research: Atmospheres*, 118(22), 12,486–12,496, doi:10.1002/2013JD020480, 2013.
- Susskind, J., Barnett, C. D. and Blaisdell, J. M.: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE TGRS*, 41, 390–409, 2003.

Chapter 8: Useful CLIMCAPS tables

Table 10: A comparison of the instruments, a priori, channel selection, information content metrics, and algorithm methods used in the CLIMCAPS and AIRS V7.

	CLIMCAPS	AIRS V7
Instruments	AIRS/AMSU (Aqua) CrIS/ATMS (SNPP, JPSS-1)	AIRS/AMSU (Aqua)
Retrieval Method	AIRS science team implementation of Optimal Estimation (OE)	AIRS science team implementation of Optimal Estimation (OE)
Retrieval variables	Temperature, water vapor, O ₃ , CO, CH ₄ , CO ₂ , HNO ₃ and others as listed in this table 10	Same as AIRS V6
Retrievals in cloudy atmospheres?	Yes, in partly cloudy atmospheres but not overcast scenes	Yes, in partly cloudy atmospheres but not overcast scenes
A-priori for T, H ₂ O and O ₃	MERRA2	Non-linear regression (neural network)
A-priori for CO, CH ₄ , HNO ₃ , SO ₂ , CO ₂ , N ₂ O	Climatology	Climatology
A-priori error propagation	2-D error covariance matrices for temperature, water vapor and ozone; 1-D error covariance matrix diagonal vector for all other trace gas species.	1-D error covariance matrix diagonal vector for all retrieval variables
Infrared spectral channels	Channel subsets are selected for each retrieval variable used in OE	All IR channels are used in non-linear regression Channel subsets are selected for each retrieval variable used in OE
Latency	1 month delay due to dependence on MERRA-2	Near real-time; No dependence on reanalysis product
Averaging Kernels	Full 2-D matrix of averaging kernels available in Level 2 product for each CLIMCAPS footprint for temperature, water vapor, O ₃ , CO, CH ₄ , CO ₂ and HNO ₃	Full 2-D matrix of averaging kernels available in Level 2 product.
Cloud Clearing	MERRA-2 is the a-priori of the cloud-free state; No dependence on AMSU or ATMS channels; Cloud cleared infrared radiances are derived in a single step for use in subsequent retrievals.	Cloud-free state is derived from non-linear regression retrieval; Iterate through two regression steps before deriving cloud cleared infrared radiances.
Cloud retrievals	Cloud fraction is retrieved for each field-of-view (FOV); Cloud top pressure is retrieved on two layers	Cloud fraction and cloud top pressure are retrieved for each field-of-view (FOV)

	for the CLIMCAPS footprint (3 x 3 FOVs)	
--	---	--

Table 11: List of CLIMCAPS V2 variables directly retrieved from the cloud cleared radiances

Retrieved Variable	Units	Fieldname	Vertical grid
Temperature	Kelvin	air_temp mw_air_temp	air_pres
Water Vapor	molec/cm ²	h2o_vap_mol_lay mw_h2o_vap_mol_lay	air_pres_lay
Ozone	molec/cm ²	o3_mol_lay	air_pres_lay
Carbon Dioxide	mol/mol	co2_vmr	air_pres
Carbon Monoxide	molec/cm ²	co_mol_lay	air_pres_lay
Cloud fraction	%	cld_frac	n/a
Cloud liquid water	molec/cm ²	mw_h2o_liq_mol_lay	air_pres_lay
Cloud top pressure	Pa	for_cld_top_pres_2lay	n/a
Methane	molec/cm ²	ch4_mol_lay	air_pres_lay
Nitric Acid	molec/cm ²	hno3_mol_lay	air_pres_lay
Nitrous Oxide	molec/cm ²	n2o_mol_lay	air_pres_lay
Sulphur dioxide	molec/cm ²	so2_mol_lay	air_pres_lay
Surface emissivity	n/a	surf_ir_emis surf_mw_emis	n/a
Surface reflectivity	n/a	surf_ir_refl surf_mw_emis	n/a
Surface temperature	Kelvin	surf_temp mw_surf_temp	n/a

Table 12: List of variables derived from CLIMCAPS V2 retrieval variables (Table 1). This is not a complete list as the product file contains many derived variables. For a full list, please see the CLIMCAPS V2 product user guide.

Derived Variables	Units	Fieldname	Vertical grid
CH4 mass mixing ratio at 400 hPa	kg/kg	ch4_mmr_midtrop	n/a
Cloud top temperature	Kelvin	cld_top_temp	n/a
CO mass mixing ratio at 500 hPa	kg/kg	co_mmr_midtrop	n/a
Geopotential height	meter	gp_hgt	air_pres
Ozone mass mixing ratio	kg/kg	o3_mmr	air_pres
Relative humidity	%	rel_hum	air_pres_h2o

Specific humidity	kg/kg	spec_hum mw_spec_hum	air_pres_h2o
Total column cloud liquid water	Kg/m ²	h2o_liq_tot	n/a
Total column ozone	kg/m ²	o3_tot	n/a
Total precipitable water	kg/m ²	h2o_vap_tot mw_h2o_vap_tot	n/a

APPENDIX B: List of Figures

- Figure 1: Flow diagram of the CLIMCAPS sequential retrieval algorithm. This gives a broad overview of the main retrieval steps and their logical flow towards two final products files, CLIMCAPS retrievals (CLIMCAPS RET) and cloud cleared radiances (CLIMCAPS CCR). Note that we discuss different aspects of CLIMCAPS algorithm flow also in (Smith and Barnet, 2019, 2020). See Table 1 for a description of the acronyms and symbols used here. 1–3
- Figure 2: Information content as ‘degrees of freedom’ for CLIMCAPS-SNPP temperature (T) retrievals (**air_temp_dof**) from full spectral resolution cloud cleared CrIS radiances for an ascending orbit (13h30 local overpass time) as a global equal-angle grid on 1.5° resolution, close to single footprint size in the lower latitudes at edge of scan. We did not apply any quality control filtering (**air_temp_qc**) since the averaging kernels (**ave_kern/air_temp_ave_kern**) from which DOF is derived are unaffected by the quality of the retrieval. DOF, instead, characterizes the potential a sounding system has in retrieving a target variable (Smith and Barnet, 2020). 2–9
- Figure 3: A diagnosis of CLIMCAPS-SNPP temperature retrievals for the North Polar latitudinal zone [$>60^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile and the error bars are the standard deviation at each pressure level. [left] CLIMCAPS temperature averaging kernel matrix diagonal vector from netCDF field **ave_kern/air_temp_ave_kern** that indicates the pressure levels at which CLIMCAPS has sensitivity to the true state of T(p) in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field **air_temp** [K]. [right] CLIMCAPS retrieval error from netCDF field **air_temp_err** [K] represented here as percentage [**air_temp_err**]/[**air_temp**]*100. CLIMCAPS uses an empirical a-priori error estimate and is represented by the thick grey line. In addition, CLIMCAPS damps temperature by 20–25% with respect to MERRA-2 to improve the retrieval estimation of trace gases. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals. In calculating these mean profiles, we filtered out all retrievals where **air_temp_qc**(*,*i,j*) ≥ 1 . We plot these profiles using the pressure level array from **air_pres***100 in hPa units. 2–10
- Figure 4: Same as Figure 2 but for the Tropical zone [30°S to 30°N]. 2–10
- Figure 5: Empirical a-priori error covariance matrix used in CLIMCAPS V2 H₂O retrievals as described in Smith and Barnet (2019). 2–12
- Figure 6: CLIMCAPS V2 smoothing error, measurement error and retrieval error covariance matrices as described in Smith and Barnet (2019). 2–12
- Figure 7: CLIMCAPS-SNPP T(p) at 500 hPa retrievals [K] from full-spectral resolution CrIS on 18 March 2019 from the ascending SNPP orbit. We filtered out all retrievals where **air_temp_qc**(*,*i,j*) > 1 , which are shown as missing values. 2–13
- Figure 8: CLIMCAPS-SNPP degrees of freedom (DOF) for H₂O vapor at every retrieval scene from ascending orbits (01:30 PM local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about H₂O vapor at every scene. For most of the globe, CLIMCAPS has H₂O vapor DOF of ~ 1 . We used the netCDF field **h2o_vap_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome. 2–17

Figure 9: Global maps of CLIMCAPS-SNPP H₂O vapor fields in the lower troposphere around 850 hPa for (left) H₂O vapor column density [molec/cm²] and (right) relative humidity [%]. H₂O column density is a retrieved variable and available in the netCDF file as **mol_lay/h2o_vap_mol_lay** on 100 pressure layers (**air_press_lay**). We selected values from layer 90 (839.98 hPa) and filtered out all retrievals where **mol_lay/h2o_vap_mol_lay_qc** > 1. Relative humidity is derived from **mol_lay/h2o_vap_mol_lay** and available in the netCDF file as **rel_hum** on 66 pressure layers (**air_press_h2o**). We selected values from layer 56 (852.79 hPa) and filtered out all retrievals where **rel_hum_qc** > 1. 2–18

Figure 10: A diagnosis of CLIMCAPS-SNPP H₂O vapor retrievals in the Tropics [$>30^{\circ}\text{S}$, $<30^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile with error bars defined by the standard deviation at each pressure layer. [left] CLIMCAPS H₂O vapor averaging kernel matrix diagonal vector from netCDF field **ave_kern/h2o_vap_ave_kern** that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of H₂O vapor in the atmosphere. [middle] CLIMCAPS H₂O vapor profile retrieval from netCDF field **mol_lay/h2o_vap_mol_lay** [molec/cm²]. [right] CLIMCAPS retrieval error from netCDF field **mol_lay/h2o_vap_mol_lay_err** [molec/cm²] represented here as percentage [**mol_lay/h2o_vap_mol_lay_err**]/[**mol_lay/h2o_vap_mol_lay**]*100. CLIMCAPS uses an empirical a-priori error estimate and is represented by the thick grey line. In calculating these mean profiles, we filtered out all retrievals where **mol_lay/h2o_vap_mol_lay_qc(i,j)** > 1. We plot these profiles using the pressure layer array called **air_pres_lay**. 2–19

Figure 11: Same as Figure 3 but for the North Polar zone [$>60^{\circ}\text{N}$] 2–20

Figure 12: Empirical a-priori error covariance matrix used in CLIMCAPS V2 H₂O retrievals as described in Smith and Barnett (2019). 2–22

Figure 13: CLIMCAPS V2 smoothing error, measurement error and retrieval error covariance matrices as described in Smith and Barnett (2019). 2–23

Figure 14: A research flight on September 18, 2019 shows (a) along a flight path over Hurricane Jerry the relative humidity profiles from (b) dropsondes released from a Gulfstream-IV “Hurricane Hunter” aircraft and (c) CLIMCAPS-SNPP H₂O retrievals as relative humidity. The solid line in (a) represents the flight path and the numbers are used to identify the location of the dropsonde profiles in (b) and (c). The colored dots indicate the center location of the CLIMCAPS footprint and if the retrieval passed (green) or failed (red). No averaging kernel convolution has been applied to the radiosonde data. However, for a quantitative comparison, we recommend that users apply this procedure. 2–24

Figure 15: **(a)** Simulated CrIS spectra for the longwave (648.75–1096.25 cm⁻¹), mid-wave (1208.75–1751.25) and shortwave (2153.75–2551.25) bands using SARTA (Strow et al., 2003) with atmospheric state defined by the first CLIMCAPS V2 retrieval (scanline=1, footprint=1) of granule 104 on 1 April 2018 from an ascending orbit (13h30 local overpass time). SARTA simulates CrIS spectra in radiance units [mW/m²/steradian/cm⁻¹], which we converted to brightness temperature [K] at scene temperature. **(b)** Absolute values of Brightness Temperature differences (dBT) to illustrate absorption features for temperature (T), ozone (O₃) and water (H₂O) vapor given a (blue) dT = 0.5 K perturbation in tropospheric T, 110–1100 hPa, (green) dT = 0.5 K perturbation in stratospheric T, 0–100 hPa, (red) dO₃ = 5% stratospheric O₃ perturbation, 0–100 hPa, and (gold) dH₂O = 2.5% tropospheric H₂O perturbation, 100–1100 hPa. Dots below the zero line indicate the CLIMCAPS retrieval

channel sets for O₃ (red) as well as the first (grey) and second (black) pass of T retrievals...2–28

Figure 16: Same as Figure 1b but for O₃ kernels with 5% perturbation in the (blue) stratosphere [0–100 hPa] and (red) troposphere [200–1100 hPa].2–29

Figure 17: Flow diagram of CLIMCAPS sequential retrieval algorithm with a focus on O₃ along three steps as discussed in text below. The full algorithm flow diagram is available elsewhere and we discussed different aspects of algorithm flow in Figures 1 and 2 of (Smith and Barnett, 2019, 2020), respectively. Output from retrieval steps defines **air_temp** and the **mol_lay** subgroup in the netCDF product file. Output from the magenta step defines the **mw** subgroup.2–30

Figure 18: CLIMCAPS-SNPP degrees of freedom (DOF) for O₃ from full-spectral resolution CrIS at every retrieval scene from ascending orbits (13h30 local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about O₃ at every scene. For most of the globe, CLIMCAPS has O₃ $1.5 < \text{DOF} < 2.5$. We used the netCDF field **o3_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome.2–32

Figure 19: A diagnosis of CLIMCAPS-SNPP O₃ retrievals in (a) North polar zone [$>60^\circ\text{N}$], (b) North mid-latitude [$>30^\circ\text{N}$, $<60^\circ\text{N}$] and (c) Tropics [$>30^\circ\text{S}$, $<30^\circ\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile with error bars defined by the standard deviation at each pressure layer. [left column] CLIMCAPS O₃ averaging kernel matrix diagonal vector from netCDF field **ave_kern/o3_ave_kern** that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of O₃ in the atmosphere. [middle column] CLIMCAPS O₃ mean profile retrieval from netCDF field **mol_lay/o3_mol_lay** [molec/cm²]. [right column] CLIMCAPS retrieval error from netCDF field **mol_lay/o3_mol_lay_err** [molec/cm²] represented here as percentage [**mol_lay/o3_mol_lay_err**]/[**mol_lay/o3_mol_lay**]*100. CLIMCAPS uses an empirical a-priori error estimate, which is represented here by the thick grey line. In calculating these mean profiles, we filtered out all retrievals where **mol_lay/o3_mol_lay_qc**(i,j) > 1 . We plot these profiles using the pressure layer array called **air_pres_lay**.2–33

Figure 20: Global maps of daytime (ascending orbit; 13h30 local overpass time) CLIMCAPS-SNPP total column O₃ in Dobson Units [DU]. We generated this map using netCDF field **o3_tot** [kg/m²] multiplied by $4.67\text{e}+04$ to convert to DU. We filtered out all values where **o3_tot_qc** > 12–34

Figure 21: Same as Figure 6 but for (a) stratospheric layer from 10.2–68.8 hPa, and (b) upper tropospheric layer from 2016.4–487.2 hPa. For this plot, we used O₃ from the netCDF field **mol_lay/o3_mol_lay** on 100 pressure layers (**air_press_lay**) and summed all values for (a) across layers 4 to 39 and for (b) across layers 56 to 75. We removed all retrievals where **mol_lay/o3_mol_lay_qc** > 12–35

Figure 22: Empirical a-priori error covariance matrix used in CLIMCAPS V2 ozone retrievals. We derived this error matrix off-line from an ensemble of co-located ECMWF and MERRA2 profiles as the covariance, $\delta\text{O}_3\delta\text{O}_3^T$, of $[\text{ECMWF} - \text{MERRA2}]/[\text{ECMWF}]$ to characterize the error as percentage (see section 2.2.4 of Smith and Barnett, 2019). The colorbar is in log-scale to enhance off-diagonal features.2–36

- Figure 23: CLIMCAPS-SNPP degrees of freedom (DOF) for CO at every retrieval scene from ascending orbits (01:30 PM local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about CO at every scene. For most of the globe, CLIMCAPS has CO DOF of ~ 1 . We used the netCDF field **co_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome. 2-39
- Figure 24: Global map of CLIMCAPS-SNPP retrieved CO tropospheric column density. CLIMCAPS retrieves CO on 100 pressure layers as column density, or number of molecules per cm^2 . Here we integrated all retrieved layers between 200–700 hPa to given an estimate of the mid-tropospheric CO load. We used the netCDF field **mol_lay/co_mol_lay**, integrated each profile into a tropospheric column density and filtered out all retrievals where their corresponding **aux/ispare_2** value was equal to one. 2-40
- Figure 25: A diagnosis of CLIMCAPS-SNPP CO retrievals for the North Polar latitudinal zone [$>60^\circ\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile and error bars the standard deviation at each pressure layer. [left] CLIMCAPS CO averaging kernel matrix diagonal vector from netCDF field **ave_kern/co_ave_kern** that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of CO in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field **mol_lay/co_mol_lay** [molec/cm^2]. [right] CLIMCAPS retrieval error from netCDF field **mol_lay/co_mol_lay_err** [molec/cm^2] represented here as percentage $[\text{mol_lay/co_mol_lay_err}]/[\text{mol_lay/co_mol_lay}]*100$. CLIMCAPS uses a CO a-priori error of 40% as represented by the thick grey line. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals, In calculating these mean profiles, we filtered out all retrievals where **aux/ispare_2**(i,j) = 1. We plot these profiles using the pressure layer array called **air_pres_lay**. 2-41
- Figure 26: Same as Figure 3 but for the Tropical zone [30°S to 30°N]. 2-42
- Figure 27: CLIMCAPS CO a-priori for the Northern and Southern Hemispheres for twelve months to capture seasonal variability in the background state. Month 1 is January, month 2 February and so on. 2-44
- Figure 28: Seasonality of the profiles for the Northern Hemisphere (left) and the southern hemisphere (right) using the time interpolation scheme outlined in this section for the year 2016. 2-44
- Figure 29: Sample time interpolation for CO first guess. The climatology dates are indicated by vertical lines, the star is the retrieval time, and w_i is the temporal weighting of each monthly climatology. If the retrieval is in the first half of month m , then the profiles for $m-1$ and m are used; if in the second half, then m and $m+1$ are used. 2-45
- Figure 30: The CLIMCAPS CO a-priori over a fixed date (April 1, 2016) to demonstrate how latitude weighting varies by hemisphere. Poleward of 30°N and 30°S contain a constant profile, and a linear interpolation is performed between 30°S and 30°N 2-46
- Figure 31: Simulated CrIS spectra using SARTA (Strow et al., 2003). (a) Longwave band [$648.75 - 1096.25 \text{ cm}^{-1}$] of top of atmosphere CrIS spectrum at full spectral resolution (FSR), given CLIMCAP retrieval scene (1,1) from granule 104 on 1 April 2018. (b) Five spectral signatures or Kernel functions to illustrate channel sensitivity to a change in the target variable as dBT/dx according to Eq. 1. They are, (red) 1% perturbation in column CO_2

[0 – 1100 hPa], (blue) 0.1 K perturbation in mid-tropospheric temperature [200 – 700 hPa], (grey) 1% perturbation for lower tropospheric H₂O [850 – 1100 hPa], (green) 1% perturbation to stratospheric O₃ [0 – 100 hPa], and (yellow) 1% perturbation to mid-tropospheric N₂O.....2-51

Figure 32: Same as Figure 31 but for the full spectral resolution (FSR) CrIS shortwave band [2153.75 – 2551.25 cm⁻¹]2-52

Figure 33: Information content as ‘degrees of freedom’ for CLIMCAPS-SNPP CO₂ retrievals (**co2_dof**) from full spectral resolution cloud cleared CrIS radiances for an ascending orbit (13h30 local overpass time) as a global equal-angle grid on (a) 1.5° resolution, close to single footprint size in the lower latitudes at edge of scan, and (b) 12° resolution as spatial aggregates (multiple values in the same grid cell were simply averaged). We did not apply any quality control filtering (**co2_vmr_qc**) since the averaging kernels (**ave_kern/co2_ave_kern**) from which DOF is derived are unaffected by the quality of the retrieval. DOF, instead, characterizes the potential a sounding system has in retrieving a target variable (Smith and Barnett, 2020). Results presented here are for CLIMCAPS-SNPP CO₂ retrievals on 1 April 2016.2-54

Figure 34: Same as Figure 31, but for (red) 1% perturbation in stratospheric CO₂ [0–100 hPa], (blue) 1% perturbation in mid-tropospheric CO₂ [200–700 hPa] and (gray) 1% perturbation in lower-tropospheric CO₂ [700–1100 hPa].2-55

Figure 35: CLIMCAPS-SNPP CO₂ retrievals as column mean mixing ratio [ppm] for all ascending orbits (13h30 local overpass time) on 1 April 2016 aggregated to (a) 3-degree equal-angle grid and, (b) 12-degree equal angle grid.....2-58

Figure 36: CLIMCAPS-SNPP CO₂ retrievals as column mean mixing ratio [ppm] averaged across all orbits (ascending and descending) from 1–30 April 2016 aggregated to (a) 3-degree equal-angle grid and, (b) 12-degree equal angle grid.....2-59

Figure 37: (a) SARTA simulated full spectral resolution CrIS Mid-Wave band (1209.75 – 1751.25 cm⁻¹) using a CLIMCAPS sounding retrieval as atmospheric state; specifically, the first retrieval (scanline=1, footprint=1) of granule number 104 on 1 April 2018 on an ascending orbit (13h30 local overpass time). The CrIS spectrum was simulated in radiance units [mW/m²/steradian/cm⁻¹] and converted to brightness temperature [K] using scene temperature. (b) Absolute values of Brightness Temperature (BT) difference to illustrate absorption features for methane (CH₄) and water (H₂O) vapor given a 1% perturbation across different parts of the atmosphere. Red: 1% CH₄ perturbation for mid-troposphere, 200–700h hPa. Blue: 1% CH₄ perturbation for lower troposphere, 700-1100 hPa. Cyan: 1% H₂O perturbation in mid-troposphere, 200–700 hPa. Green: 1% H₂O perturbation in lower troposphere, 700-1100 hPa.2-63

Figure 38: Same as Figure 37 but for absolute values of Brightness Temperature (BT) differences of CH₄ and H₂O across a single tropospheric layer (200–1100 hPa), with varying perturbation values for H₂O. Red: 1% CH₄ perturbation. Blue: 1% H₂O perturbation. Cyan: 5% H₂O perturbation. Green: 10% H₂O perturbation.2-64

Figure 39: CLIMCAPS-SNPP CH₄ column density retrievals [molec/cm²] from full-spectral resolution CrIS integrated over mid- to upper tropospheric pressure layers between 200 and 700 hPa. The CLIMCAPS CH₄ column densities were gridded and averaged on a 1.5-degree global grid to closely resemble variation at native product resolution in the Level 2 files.

Note that the color scale is amplified to highlight spatial patterns **mol_lay/ch4_mol_lay** field, and quality control metrics in **mol_lay/ch4_mol_lay_qc**. We integrated and averaged all values where **mol_lay/ch4_mol_lay_qc** = 0 (i.e., best quality). Much of the variability in CH₄ visible here can be attributed to variability in H₂O vapor, another gas species radiatively active in the 1300cm⁻¹ spectral range from which CLIMCAPS select its CH₄ channels for retrieval. Compared to CH₄, H₂O has a stronger spectral signal which complicates the separation of these two gases during retrieval. 2–65

Figure 40: CLIMCAPS-SNPP degrees of freedom (DOF) for CH₄ from full-spectral resolution CrIS at every retrieval scene from ascending orbits (13h30 local overpass time) on 1 April 2016. DOF is an information content metric and quantifies how many pieces of information (or distinct vertical layers) CLIMCAPS can retrieve about CO at every scene. For most of the globe, CLIMCAPS has CH₄ DOF < 1. We used the netCDF field **ch4_dof** and did not apply any quality filtering since DOF is not affected by retrieval outcome. 2–66

Figure 41: A diagnosis of CLIMCAPS-SNPP CH₄ retrievals for the North Polar latitudinal zone [$>60^{\circ}\text{N}$] on 1 April 2016. Each solid line represents the mean zonal profile and error bars the standard deviation at each pressure layer. [left] CLIMCAPS CH₄ averaging kernel matrix diagonal vector from netCDF field **ave_kern/ch4_ave_kern** that indicates the pressure layers at which CLIMCAPS has sensitivity to the true state of CH₄ in the atmosphere. [middle] CLIMCAPS CO profile retrieval from netCDF field **mol_lay/ch4_mol_lay** [molec/cm²]. [right] CLIMCAPS retrieval error from netCDF field **mol_lay/ch4_mol_lay_err** [molec/cm²] represented here as percentage $[\text{mol_lay/ch4_mol_lay_err}]/[\text{mol_lay/ch4_mol_lay}]*100$. CLIMCAPS uses a CH₄ a-priori error of 5% as represented by the thick grey line in panel on the right. A Bayesian Optimal Estimation retrieval system (like CLIMCAPS) typically reduces the a-priori error in all successful retrievals. In calculating these mean profiles, we filtered out all retrievals where **mol_lay/ch4_mol_lay_qc**(*,*i,j*) ≥ 1 . We plot these profiles using the pressure layer array from **air_pres_lay***100 in hPa units. 2–67

Figure 42: Same as Figure 41 but for the Tropical zone [30°S to 30°N]. 2–68

Figure 43: Same as Figure 39 but with values spatially aggregated over a 4° equal-angle grid. 2–69

Figure 44: Same as Figure 39 but with values temporally aggregated over a 4° equal-angle grid for a month of retrievals, ascending (13h30 local overpass time) and descending orbits (01h30 local overpass time), 1–30 April 2016. 2–70

Figure 45: CLIMCAPS CH₄ a-priori mixing ratio [ppb] profiles calculated along each latitude from North to South. The CLIMCAPS CH₄ a-priori is calculated from a set of coefficients and varies with pressure [hPa] and latitude. These coefficients were developed by Xiong et al.(2008, 2013). 2–71

Figure 46: An illustration showing the differences between CLIMCAPS pressure levels (P_{lev}) [**air_pres**] and pressure layers (P_{lay}) [**air_pres_nsurf**] for an index n , where $n \in 1 \dots 100$. Temperature is retrieved on levels while trace gasses are retrieved on pressure layers. The mathematical conversion from levels to layers is shown in equation 6. 3–77

Figure 47: An illustration of the two types of boundary layer conditions that result when retrieving atmospheric variables on a standard pressure grid, P_{obs} , which for CLIMCAPS is the pressure levels profile, **air_pres**. CLIMCAPS uses surface pressure, P_s , from MERRA2.

Here the subscript n denotes ranges from 1 to 100, and represents the index at which the CLIMCAPS retrieval is reported at. Unless the retrieval is performed over ocean $P_{\text{obs}(n)}$ is always less than $P_{\text{obs}(\text{plev})}$ with $n < \text{plev}$, $\text{plev} = 100$, and $P_{\text{obs}(\text{plev})} = 1100$ mb. Scenario 1 requires a narrowing of the boundary layer with interpolation of the retrieved value from $P_{\text{obs}(n-1)}$ to P_s . Scenario 2 requires a broadening of the boundary layer with extrapolation of the retrieved value from $P_{\text{obs}(n-1)}$ to P_s . The brown dotted line indicates the width of the adjusted boundary layer, which always exceeds 5 mb and varies horizontally with surface topography. 3-79

Figure 48: The CLIMCAPS retrieval footprint is the field-of-regard (FOR; grey dashed circles) that consists of 3 x 3 instrument fields-of-view (FOV; black solid circles). CLIMCAPS aggregates the 9 FOVs into a single spectrum from which it then retrieves a set of atmospheric profile variables. Cloud fraction is the only variable that CLIMCAPS retrieves for each FOV (9 per FOR); all other retrieval variables represent conditions within the FOR (~50km at nadir, ~150km at edge-of-scan). Here we illustrate four typical cloudy scenarios encountered by CLIMCAPS: (a) partly cloudy FOR where all FOVs have cloud fraction > 0.0 (i.e., not clear) but no two FOVs has the same cloud fraction, (b) partly cloudy FOR where some FOVs have no clouds (cloud fraction = 0), (c) partly cloudy where each FOV has the exact same cloud fraction (no contrast), and (d) overcast FOR where all FOVs have cloud fraction = 1.0. Cloud clearing is accurate (i.e., small brightness temperature residuals with low **etarej** values) in (top row) spatial *heterogeneous* retrieval footprints, but fails (i.e., large brightness temperature residuals with high **etarej** values) in (bottom row) spatial *homogeneous* scenes. 4-84

Figure 49: Scene-dependence of CLIMCAPS-Aqua averaging kernels for coincident (top row) temperature (air_temp) and (bottom row) water vapor (h2o_vap) retrievals at five scenes (left to right) on 1 July 2018 from Granule 60 with 13h30 local overpass time. The latitude/longitude coordinates are listed at the top of each figure. Averaging kernels quantify and characterize the signal-to-noise ratio of an observing system and are affected by the scene-dependent effects (e.g., temperature lapse rate, amount of gas molecules, surface emissivity and cloud uncertainty) as much as the measurement characteristics (e.g., spectral resolution, instrument calibration and noise). CLIMCAPS retrieves air_temp and h2o_vap sequentially each with a unique subset of channels, which means that the variation in these averaging kernels are independent of each other. 5-91

Figure 50: Same as Figure 48 but for CLIMCAPS-JPSS1 averaging kernels from Granule 97 on 1 July 2018 and 13h30 local overpass time. 5-92

Figure 51: The mean (solid line) and standard deviation (error bars) of averaging kernel matrix diagonal vectors in the tropics (30°S to 30°N) on 1 July 2018 and from (top) CLIMCAPS-Aqua and (bottom) CLIMCAPS-JPSS1, both ascending orbits. The error bars indicate the degree to which the averaging kernel diagonal vectors vary spatially across the latitudinal zonal. CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020). 5-93

Figure 52: Same as Figure 50 but for southern polar zone (90°S to 60°S). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020). 5-94

- Figure 53: Same as Figure 50 but for southern mid-latitude zone (60°S to 30°S). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020). 5–94
- Figure 54: Same as Figure 50 but for northern polar zone (60°N to 90°N). CLIMCAPS AKs within the northern mid-latitude zone was published and discussed in (Smith and Barnett 2020). 5–95
- Figure 55: Spatial variation of CLIMCAPS information content for different retrieval variables and different space-based instruments from their ascending orbits (13h30 local overpass time) on 15 December 2018. (Top row) A global view of CLIMCAPS averaging kernel values at ~500 hPa for H₂O vapor retrievals from (top left) AIRS/AMSU on Aqua, and (top right) CrIS/ATMS on JPSS-1. (Bottom row) A global view of CLIMCAPS averaging kernel values at ~700 hPa for temperature retrievals from (bottom left) AIRS/AMSU on Aqua, and (bottom right) CrIS/ATMS on JPSS-1. 5–96
- Figure 56: Spatial and vertical variation of CLIMCAPS-JPSS1 ascending orbit (13h30 local overpass time) information content for ozone at different pressure values from the stratosphere to upper troposphere on 15 December 2018. The panels are CLIMCAPS-JPSS1 ozone averaging kernel diagonal values at (top left) 60 hPa (top right) 86 hPa, (bottom left) 174 hPa, and (bottom right) 253 hPa. 5–97
- Figure 57: Trapezoid state functions for air_temp, h2o_vap, O₃, CO, CH₄, CO₂ and HNO₃, which define the pressure levels on which CLIMCAPS Jacobians (weighting functions) and thus their averaging kernels are calculated. 5–99
- Figure 58: Demonstrating the value of convolving methods. Originally presented by Chris Barnett in: http://www.weatherchaos.umd.edu/group_log/data/y0910/091019_weatherchaos_barnet.pdf 5–101
- Figure 59: Fraction of retrievals with a given QC flag for CLIMCAPS-SNPP T(p) retrievals (**air_temp_qc**) from full spectral resolution cloud cleared CrIS radiances. The results are shown for five latitudinal zones and for the full global retrieval. ‘Best’ quality data has a QC flag value of 0, ‘best+good’ has a value of 0 or 1, and rejected data has a value of 2. Note that while each variable has a corresponding *_qc field, these values do not vary between CLIMCAPS state and derived variables. The example presented here is from CLIMCAPS-SNPP retrievals on 1 April 2016. 6–110
- Figure 60: Example of a footprint-level QC approach used in the NUCAPS algorithm by the National Weather Service (NWS). Green retrievals indicate the IR+MW step successfully passed, yellow retrievals where the MW-only step passed while the IR+MW step failed, and red indicates both steps were rejected. The example is taken from a screen capture from AWIPS-II of NUCAPS from a NOAA-20 overpass of convection at 19:45 UTC on April 24, 2019. 6–111
- Figure 61: Fraction of retrievals with a given QC flag for CLIMCAPS-SNPP T(p) retrievals (**aux/ispare_2**) from full spectral resolution cloud cleared CrIS radiances. The results are shown for five latitudinal zones and for the full global retrieval. ‘MW+IR Pass’ data has a bit flag of zero, ‘MW-only Pass’ has a value of 1, and ‘Reject’ has a value of 9. Results presented here are for CLIMCAPS-SNPP retrievals on 1 April 2016. 6–112

Figure 62: CLIMCAPS V2 channel sets for AIRS on Aqua in the (top) long-wave infrared (IR) band, (middle) mid-wave IR band and (bottom) short-wave IR band. Each item on the y-axis represents a CLIMCAPS retrieval variable with vertical lines indicating the channel wavenumbers for each variable. From top to bottom there is, Nitrous Oxide (N₂O), Carbon Dioxide (CO₂), Sulphur Dioxide (SO₂), Nitric Acid (HNO₃), Methane (CH₄), Carbon Monoxide (CO), atmospheric temperature (Air T) first and second passes (see flow diagram), Ozone (O₃), Water Vapor (H₂O), cloud clearing (CC), cloud top pressure (CTP) and cloud top fraction (CTF), Earth surface skin temperature (T_{skin}) and Reflectivity (Refl). 7-118

Figure 63: Same as Figure 62 but for CrIS full spectral resolution (FSR) on SNPP and JPSS1 (NOAA20). 7-119

Figure 64: Same as Figure 62 but for CrIS normal spectral resolution (NSR) on SNPP. 7-120

Figure 65: Same as Figures 62 through 64, but instead summarized according to retrieval variables with the y-axis items the three instrument configurations. On the right-hand y-axis of each panel, we list the total number of channels for selected for each instrument and target variable. 7-121

Figure 66: Wavenumbers of channels used in the retrieval of only one atmospheric variable, i.e., the set of unique spectral channels for the three infrared bands, (top) long-wave, (middle) mid-wave and (bottom) short-wave. We depict these for (grey) CrIS NSR, which has 440 unique channels, (blue) CrIS FSR with 553 unique channels and (red) AIRS with 541 unique channels. On the right-hand side of each panel, we list the total number of unique channels for each instrument and each spectral band. 7-122

Figure 67: Same as Figure 66 but depicting the channels shared by two or more retrieval variables for each instrument and IR spectral band. CrIS NSR has a total of 52 channels used in multiple CLIMCAPS retrievals, CrIS 146, and AIRS 126. 7-123

Figure 68: A break-down of the CLIMCAPS spectral channels used in two or more retrievals from AIRS measurements. The number and target of the shared channels vary in the (top) long-wave, (middle) mid-wave and (bottom) short-wave spectral bands. 7-124

Figure 69: Same as Figure 64 but for CrIS full spectral resolution (FSR). 7-125

APPENDIX B: List of Tables

Table 1: List of acronyms and symbols used in Figure 1.	1-4
Table 2: Channel selection from AMSU and ATMS for each of the two CLIMCAPS retrieval stages, microwave only (MW-only), and a combined MW and IR (MW+IR) retrieval. See CLIMCAPS flow diagram for methodology and IR channel selection chapter for the IR channel subsets.	2-8
Table 3: Channels selected from AMSU and ATMS for each of the two CLIMCAPS retrieval stages, microwave (MW) only, and a combined MW and IR retrieval. See CLIMCAPS flow diagram for methodology and IR channel selection chapter for the IR channel subsets. ...	2-17
Table 4: NUCAPS uses a linear regression for temperature T(p) retrievals, a global climatology for CO ₂ that scales linearly over time, is available in near real-time (NRT), has no averaging kernels in its product file and no minor constituent detection flags. The AIRS V6 system uses a non-linear regression (or neural net NN) for T(p) retrievals, does not have a distinct CO ₂ a-priori, is available in NRT with partial averaging kernels. CLIMCAPS uses MERRA2 as T(p) a-priori, has the same CO ₂ climatology a-priori as NUCAPS, is not available in NRT but instead lags by a month (due to MERRA2), has the full averaging kernel matrices available at each retrieval footprint for each retrieval variable and has detection flags for minor constituent.	2-57
Table 5: Summary of CLIMCAPS averaging kernel variables in the product file. All fields listed here are from the ave_kern/ group inside the netCDF file. P _{top} is the pressure layer value [hPa] of the top-most Trapezoid state function, and P _{bot} , the pressure layer value of the bottom-most state function. We read P _{top} and P _{bot} from the corresponding *_ func_pres fields.	5-98
Table 6: A summary of the output created by <code>slb2fin.pro</code> for a single trapezoid state function. Here we use CLIMCAPS O ₃ as example and list the content of [ave_kern/o3_func_idx] in column 2. The standard 100-level pressure array [air_pres] in hPa units is listed in column 3. All other column titles refer to the variable names as used in <code>slb2fin.pro</code> and <code>calc_finv.pro</code> . Column 4 ('func_ampl') is defined in <code>calc_finv.pro</code> and used as one of the input variables to <code>slb2fin.pro</code> , and indicates that the fourth O ₃ trapezoid state function should be calculated (row 4 of column 4 = 1.0). The last column ('slope') defines the shape of the state function.	5-102
Table 7: Description of quality control values and their appropriate use.	6-108
Table 8: Overview of the two hyperspectral infrared (IR) instruments on low-Earth orbiting satellite platforms that CLIMCAPS is configured for. CrIS on SNPP was at first transmitted at nominal spectral resolution (NSR) but later switched to full spectral resolution (FSR) so we list statistics for both. These instruments measure the top of atmosphere radiance with spatial footprints known as (FOV) and CLIMCAPS retrieves soundings from every 3 x 3 array of FOVs known as the Field of Regard (FOR). Instrument noise is given in brightness temperature units as the noise equivalent delta temperature (NEDT) for a scene at 250K. This table was originally published in (Smith and Barnett, 2019). Note that we exclude mention of the two CrIS duplicate 'guard' channels for each spectral band.	7-116

- Table 9: Spectral channel summary for Aqua/AIRS, Suomi-NPP CrIS normal spectral resolution (CrISNSR), Suomi-NPP CrIS full spectral resolution (CrISFSR), and JPSS-1 (or NOAA-20) CrISFSR, and eight CLIMCAPS retrieval variables. These channel totals are for apodized CrIS radiances, excluding count of the ‘guard’ channels. For each set of channels, we list the average degrees of freedom (DOF) achieved within CLIMCAPS for a global day of FORs. This table is an extract from one published in (Smith and Barnett, 2020). 7-117
- Table 10: A comparison of the instruments, a priori, channel selection, information content metrics, and algorithm methods used in the CLIMCAPS and AIRS V7. 8-127
- Table 11: List of CLIMCAPS V2 variables directly retrieved from the cloud cleared radiances 8-128
- Table 12: List of variables derived from CLIMCAPS V2 retrieval variables (Table 1). This is not a complete list as the product file contains many derived variables. For a full list, please see the CLIMCAPS V2 product user guide. 8-128